JOURNAL of Applied Economic Sciences

JAES

Journal of Applied Economic Sciences

Volume III
Issue 3(5)

Fall 2008

Publisher:

European Research Center of Managerial Studies in Business Administration

Email: jaes_secretary@yahoo.com

I.S.S.N. 1843-6110

Journal of Applied Economic Sciences

Journal of Applied Economic Science is a young economics and interdisciplinary research journal, aimed to publish articles and papers that should contribute to the development of both the theory and practice in the field of Economic Sciences.

The journal seeks to promote the best papers and researches in management, finance, accounting, marketing, informatics, decision/making theory, mathematical modelling, expert systems, decision system support, and knowledge representation. *This topic may include the fields indicated above but are not limited to these.*

Journal of Applied Economic Sciences be appeals for experienced and junior researchers, who are interested in one or more of the diverse areas covered by the journal. It is currently published quarterly with three general issues in Winter, Spring, Summer and a special one, in Fall.

The special issue contains papers selected from the **International Conference** organized by the **European Research Centre of Managerial Studies in Business Administration** and **Faculty of Financial Management Accounting Craiova** in each October of every academic year. There will prevail the papers containing case studies as well as those papers which bring something new in the field. The selection will be made achieved by:

- Conference Session Chairs
- Editorial Board Members

Journal of Applied Economic Science is indexed in CEEOL - www.ceeol.org, EBSCO www.ebsco.com and in RePEc www.repe.org.

The journal will be available on-line and will be also being distributed to several universities, research institutes and libraries in Romania and abroad. To subscribe to this journal and receive the on-line/printed version, please send a request directly to jaes_secretary@yahoo.com

Journal of Applied Economic Sciences

ISSN 1834-6110

Table of Contents

ASSESSMENT OF INCOME DISTRIBUTION AND A HYPOTHETICAL FLAT TAX REFORM IN HUNGARY

Dóra **BENEDEK**,
Central European University, Budapest, **HUNGARY**
dora.benedek@gmail.com
Orsolya **LELKES**,
European Centre for Social Welfare Policy and Research, Vienna, **AUSTRIA**
lelkes@euro.centre.org

Abstract:

The paper presents evidence on the effects of taxes and benefits on household incomes in Hungary referring to the 2006 system and a hypothetical flat tax reform. For this, a microsimulation model is used, which is based on a matched sample of an income and a consumption survey and administrative tax records. The Hungarian budget receives more revenues from VAT than from PIT. This has major implications on equity, as while PIT is progressive, VAT is regressive, imposing a higher tax burden on low-income households. We highlight the importance of tax allowances. The absolute amount of total tax allowances tends to increase with income, and the share of allowances within total incomes is around 5-7% in all income groups, except the top fifth, where it declines. Targeting is thus inadequate, and it is especially so in case of child support. Family tax allowance reaches the bottom decile only to a limited extent. This is in sharp contrast with the universal child benefit, which is well targeted to the poorest. The second part explores the likely impact of the introduction of a flat tax, where VAT and PIT rates are set at 20%, and a tax free bracket for low incomes is kept. We show that a budget neutral solution would have a largely regressive effect, where 70% of the population would lose, with a minority on the top of the distribution gaining.

Keywords: tax-benefit microsimulation, redistribution, flat tax reform, Hungary

JEL Classification: C8, D31, I38

1. Introduction

In Hungary, discussions on economic policy priorities often appear misguided. A typical argument is that less state and less redistribution is economically superior to more redistribution, which appears to be a simplistic backlash to the paternalist model of the socialist economy. Key indicators used in setting policy priorities are the ratio of spending or the ratio of revenues compared to the GDP. Yet little is known on the impacts of the structure of public spending and that of revenues, be it either short term or long term, and perhaps even less known is the interplay of the various specific instruments. Yet an "opposite" group of social scientists argue for increased social spending. Many of these arguments, however, disregard the opportunity cost of spending or the existence of budget constrains as such. A "caring state is a spending state" may be the logo of such argument. We believe that the focus of the debate needs to be fundamentally reshuffled, with more focus on empirical evidence and on microeconomic effects. We aim to promote this process with the presentation of research evidence on the redistributive impacts of the tax and benefit system as a whole. For this, we will use the national tax-benefit microsimulation model of the Ministry of Finance.

2. Motivation of the paper

Little is known on the effects of the tax-benefit system on the income distribution in Hungary. Empirical evidence, using Euromod, the European tax-benefit microsimulation model and the national models of Hungary and Poland compares the 2001 system of Hungary with those of EU15 countries and Poland [Sutherland, Levy and Paulus, (2006)]. The study finds that the distribution of original incomes is rather unequal, with a shape that is rather similar to that of the UK and Ireland, in contrast to the more equal distribution of Austria and Germany The role of the state in Hungary, however, is rather different from these Anglo-Saxon countries: due to its more generous public pension scheme, a considerable share of the benefits go to the top fifth as well, although to a much smaller extent than e.g. in France, Austria or Poland[i]. The tax system also seems to play a specific role in Hungary: taxes (including tax allowances), seem to be rather efficient in raising the incomes of those in the second to fourth deciles, in contrast to all other countries, where the role of direct taxes is smaller and varies less in the bottom half of the income distribution. Contrary to popular belief, the tax burden on the top fifth seems to be rather close to the lower end of the range in the EU15. Taxes and contributions range between 22% (Portugal) and 46% (Denmark) as a proportion of gross incomes of the top quintile, while the ratio is 27% in Hungary. These comparative figures, however, do not highlight the role of specific benefits, neither those of tax allowances, and are constrained to distributional analysis by income quintiles, rather than other demographic characteristics.

Hungary appears to be a useful case study not only for the lack of detailed national evidence so far. Inequality has increased considerably during the years following economic transformation, both in terms of material resources but also in subjective well-being [Tóth, (2005), Lelkes, (2006)]. With the joining of the European Union, the country has officially become part of the EU's decision making process, including increased monitoring and the production of social indicators. Yet little is known on many policy instruments and their social impacts, for example tax allowances, and general welfare typologies, such as the much cited one of Esping-Andersen (1990) may not be applicable or particularly helpful for a potential European harmonisation [Lelkes, (2000)].

Flat taxes seems to be particularly popular in much of Eastern Europe, and have been introduced in the Baltic states, Russia, Ukraine, Georgia, Slovakia, Serbia, Romania, Macedonia, although with a varied definition on what is actually meant by flat tax. In many of these countries, personal income tax schemes sustained a tax free bracket for those with low incomes often with specific child tax credits. In addition, a social security contribution is often added to the seemingly low income tax rates, constituting a higher total tax burden in an economic sense. The evidence on the impact of these reforms seems to be sparse, partly because it proved to be difficult to disentangle the effect of flat tax reform from other parallel policy reforms, or external macroeconomic effects, and partly, because the lack of adequate micro data hinders such analyses. In Russia, for example, where revenues boosted following the tax reform, economic growth and high oil prices are likely to have played a dominant role rather than changing tax compliance, according to the IMF.

The introduction of a flat tax scheme is a recurring theme in Hungarian politics as well, although with not much reference to the details of such reform. The proponents believe that a significant simplification of the current scheme would increase tax compliance, and therefore would ultimately enable the reduction of the tax burden. Some also regard this tax scheme fairer than the current (progressive) one. The underlying core issue, however, seems to be the level of taxes and the complexity of the tax system as such, rather than the rates themselves. We argue that a flat tax scheme is not a precondition for the achievement of these goals, as the complexity of the system does not arise from the number of rates, but rather from the number of exemptions and the complex rules of defining the tax base. We show that the introduction of a flat income tax with a budget neutral solution (on a short run) would have a largely regressive effect, where the majority of the population would lose, with a minority on the top of the distribution gaining. These adverse effects need to be born in mind when discussing policy alternatives.

First we briefly present the TÁRSZIM2005[ii] model, including the underlying dataset, then analyse the redistributive effects of the tax system and that of cash benefits of 2006. The concluding section presents the potential impact of the introduction of a flat tax system in Hungary.

3. The Hungarian microsimulation model: TÁRSZIM

In many countries microsimulation techniques have been widely used in government and the academia, so there is little need to prove the legitimacy of it. For example, in the UK the Treasury and the Department for Work and Pensions publish their results regularly during the debate of the budget. EUROMOD, the European tax-benefit model has a whole network of researchers across Europe, with a diverse set of research results[iii]. Many governments in Eastern Europe, including e.g. Estonia, Poland, Slovenia have already commissioned the construction of such a model, although interestingly often at the initiation of independent researchers, who were well aware of the benefits and methodology of the technique[iv].

Microsimulation modelling has not so far been used in Hungary to prepare economic policy decisions, despite some earlier initiatives by the CSO and TÁRKI[v]. It was only in 2004, with a supportive finance minister, and the establishment of a new research unit, that the use of the technique became possible. The tax-benefit model, named TÁRSZIM, was developed by the TÁRKI social research institute, and was commissioned by the Ministry of Finance.

3.1. Model features

The model includes personal income tax (PIT); tax allowances; indirect taxes (VAT); and the major social benefits [for details see Benedek and Lelkes, (2005) and TÁRSZIM2005 Manual]. As is the case for all such models, the parameters of all these policy instruments can be set according to actual or hypothetical scenarios, including for example the tax rates, and the entitlement criteria.

Similar to other existing tax-benefit microsimulation models, it only refers to the cash part of redistribution, thus ignores the consumption of public goods or benefits in kind, such as health care or education. The reason is that the pricing of these social services is complex on an individual level, the information on the consumption of the specific items is inadequate, and the implications of their consumption on inequality is often unclear[vi].

The model is able to simulate all central government taxes and benefits, but not local government ones. There are 3200 local governments in Hungary, with varying policies, so the impact of their practices (and potential policy changes) cannot possibly be modelled with the available dataset. Household revenues from local sources, however, constitute part of total incomes, as they are part of the income survey.

The unit of impact assessment can be the household or the individual. The model thus allows to analyse the interaction between household members. We can follow how changes in individual incomes (as result of policy changes) add up on a household level. This is of major relevance, as the unit of policy is often the individual, while incomes are shared within the household. Depending on the household composition, e.g. whether jobless people tend to be concentrated in jobless households or rather live with employed spouses, the impact of policies may vary a great deal.

TÁRSZIM is a static model, therefore it is suitable for assessing "day-after" effects, but it is not suitable for assessing behavioural responses in the area of labour market activity, and only to a very limited extent is it capable of assessing changes in consumption patterns.

3.2. The database

The database consists of three different datasets, an income and a consumption survey and an administrative data on tax records, which have been merged with probabilistic matching, based on region, age, income decile, and gender. The merged dataset thus includes information on income, consumption and taxation, beyond the essential socio-demographic information on households.[vii] The core dataset is the 2004 TÁRKI Monitor data, which contains individual demographic, labour market and household characteristics and income data of 2325 households and their members for 2003. This is supplemented by the 2003 database of the CSO (Central Statistics Office) Household Budget Survey containing detailed consumption data for about 8 thousand households. The third database is a random sample of almost 62 thousand observations, compiled by APEH (the Tax and Financial Control Administration), from the 2003 personal income tax returns.

The core dataset thus contains data from 2003. For analyses that relate to any year other than 2003 the basic settings need to be adjusted. Multipliers (adjusting for nominal growth between 2003 and 2006) are applied for the model used in this study. Multipliers regarding income items are calculated by the experts of the Ministry of Finance, while the price indices applied to consumption items are based on CSO data.

The VAT estimates of the model seem rather flat across income groups due to two types of errors. First, there was a rather high unit non-response rate in the original dataset, the CSO Household Budget Survey. The unit non-response was higher among the well-off, but unfortunately the actual extent of the problem is not revealed by the Statistical Office [CSO, (2004) pp. 29). Second, the model assumes full tax compliance, in other words, that everyone pays the tax due on each HUF spent on consumption. As a result of all distortions the model seems to underestimate total VAT by 25% and VAT payment by higher income groups even more. As a result, we expect VAT to be more progressive in reality than estimated by the model.

In the next session we present the redistribution of the 2006 tax and benefit system in Hungary.

4. Effects of the 2006 tax and benefit system on income distribution

The poverty reduction effect of benefits is compelling: if the government suspended all cash benefits, poverty increased from 13% to 33% among the working age population, according to 2003 data, using a 60% of the national median income as a poverty threshold. Among the elderly, defined as those above the age of 65, poverty would multiply: from 9% to 67%. Similar calculations can be made for specific cash benefits, such as universal child benefit, unemployment benefit [e.g. Förster-Tóth (1997, 2001)]. These calculations are appropriate for presenting the effect of benefits on specific groups of the society and can demonstrate the relative importance of certain benefits. The limitation of these analyses is that they can be done only ex post, once data is available on past years, and that they can be done only for very specific policy instruments, disregarding the interaction of these. Therefore, their role in decision-making is limited. As opposed to these methods, microsimulation can be used for a comprehensive analysis of various instruments, and can be used for testing future policy scenarios.

4.1. Parameters of the tax and benefit system[viii]

Taxes on incomes:

▪ In 2006 a two-tier personal income tax regime was in place, with the following tax rates:

0 –	1,550,000	18%
1,550,001 –		38%

▪ capital incomes are subject to 10%, 20%, 27% or 35% tax rates;

▪ the rates of employee pension contributions and health insurance contributions are 8.5% and 4%, respectively;

▪ employee tax allowance and supplementary tax allowance ensure that the minimum wage is tax exempt;

▪ family tax allowance applies to those with 3 or more children, may be shared between parents;

▪ other tax allowances refer to repayment of subsidised mortgage loans, life insurance and pension savings, adult education, charitable donations, etc.

Taxes on goods:

▪ there is a motor vehicle registration fee;

▪ the normal VAT rate is 20%, with a preferential rate of 15%. Certain products and services are subject to an even lower VAT rate;

▪ excise duty is levied on alcoholic beverages, tobacco products and fuel;

Major benefits:

▪ means tested: regular social assistance and housing benefit

▪ universal: child benefit and other family benefits (maternity benefit, etc.)

▪ gas-price subsidy

▪ other, insurance based benefits: pension, unemployment benefits, etc.

4.2. Redistribution in the 2006 tax and benefit system

First we analyse the redistribution effects of the tax and benefits system on the total population, that is, on all households. Besides taxpayers this category includes children, pensioners, the unemployed and other economically inactive groups. Figure 1 shows the tax liabilities and the social benefits of households (including gas price subsidy) as a percentage of disposable income. Substantial redistribution can be observed: tax payment accounts for a smaller percentage of disposable income than social benefits in the lower half of the distribution, i.e. the state collects smaller amounts from the first five deciles in the form of direct taxes than it returns to them in the form of benefits. Yet from the seventh decile upward households pay more to the budget than they receive as benefits. However if social security contributions are also taken into account, which finance mostly insurance based benefits, then the middle income groups are net financers of the system. It is also worth noting that although the top 3 deciles receive a substantial part of the benefits, it only counts for 2-5% of their disposable incomes (Figure 2).

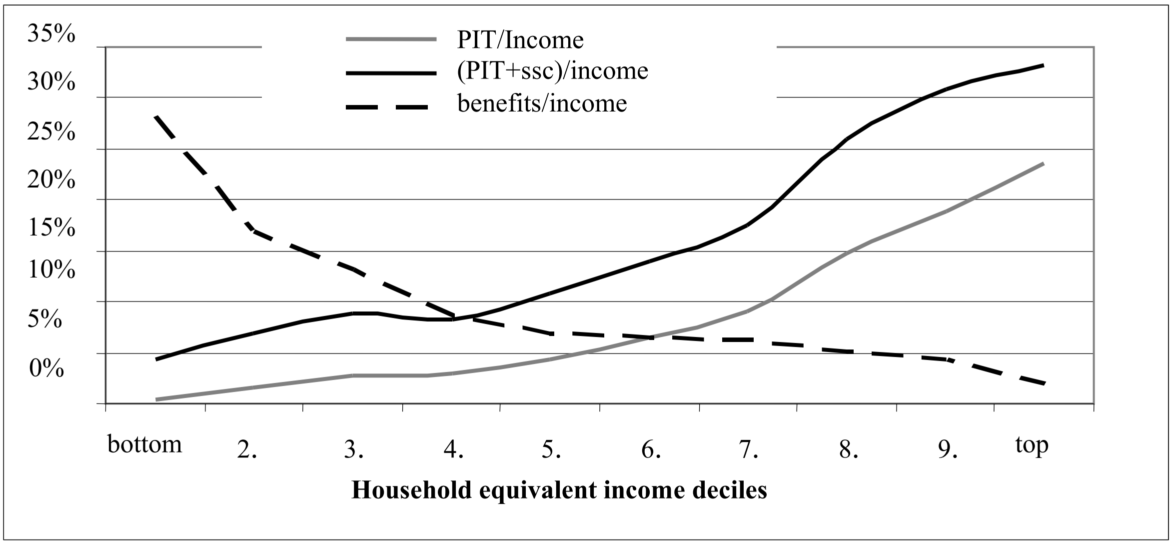


Figure 1 . Income taxes and cash benefits across income groups. Taxes (PIT), social security contributions (ssc) and benefits as a ratio of disposable income, %, households

Source: own calculations with TÁRSZIM2005

Note: benefits = universal child benefit + maternity benefits + gas price subsidy + regular social assistance + housing benefit

Figure 2 shows that a great proportion of benefits (plus family tax allowance) is allocated to the richer half of the population: the top five deciles get over 40% of the benefits. The bottom two decile groups get no more than 30% of the total spending. This may not be a problem per se, given that the welfare system's main function is not poverty alleviation per se, but also the smoothing of individual's incomes over the life cycle. Nevertheless, this ratio is not known and worth to be noted. Note also, that this figure does not say anything on the poverty alleviation effect of these transfers, nor does it include local means-tested social assistance. A final word on the results of the table: the decrease at the middle of the distribution is due to the fact that a high proportion of pensioners can be found here, who receive only a negligible share of the family benefits.

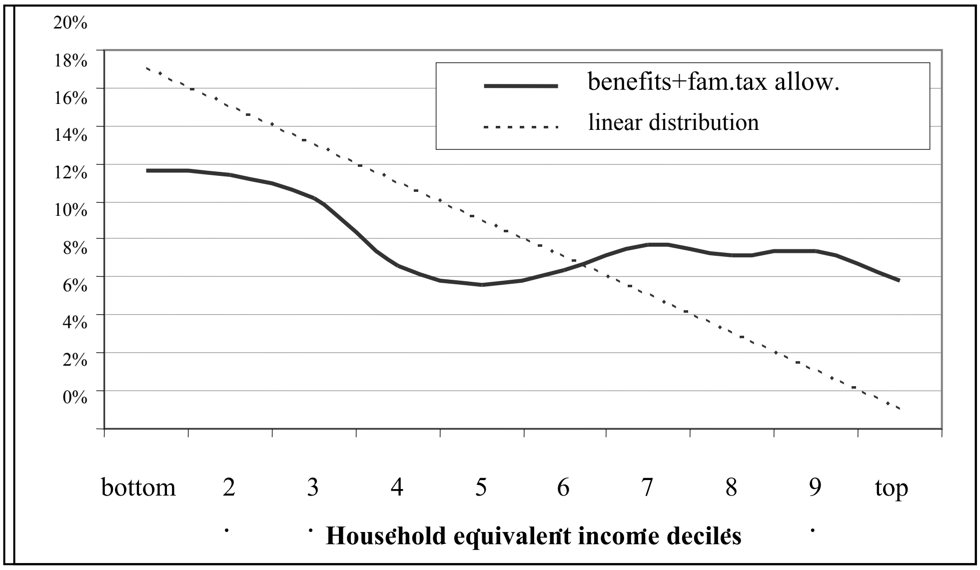


Figure 2: Where does budget spending go?
Distribution of the total sum of benefits and family tax allowance among deciles of households

Source: own calculations with TÁRSZIM2005

There is a great difference between the „intended” and actual progressivity of the PIT system, due to the impact of tax allowances and tax credits. This part of the analysis was conducted on the group of taxpayers, who make up less than 50% of the total population. In Figure 3 the per capita gross income and tax liability are arranged by net income, along with the implicit tax rate calculated as a ratio of the two. The difference between the two lines highlights the redistribution effect of tax allowances. The dashed line, the ratio of calculated tax to gross income, shows the rate of tax that *would be* borne by individuals in the various deciles *without the tax allowances.* This varies between 17% and 25% for the various groups, which is in line with the tax scheme. The implicit tax rate, indicated by the continuous line on the graph, which shows the ratio of *actual tax liability* to gross income taking tax allowances into account, is much more progressive and is more favourable to the less affluent groups of society. This rate is below 5% in the lower deciles, and then gradually rises along the income, until finally, in the top decile, it is up to 25%, which is almost as high as the rate calculated without the tax allowances. This suggests that the progressivity of the PIT system is rather weak when only the impact of tax brackets is considered, but becomes rather strong due to the effect of tax allowances.

The slight increase in the 2nd and 3rd decile groups is due to the fact that most of the self-employed belong here based on their declared income and they are not eligible for the most substantial tax allowance, the employee tax allowance.

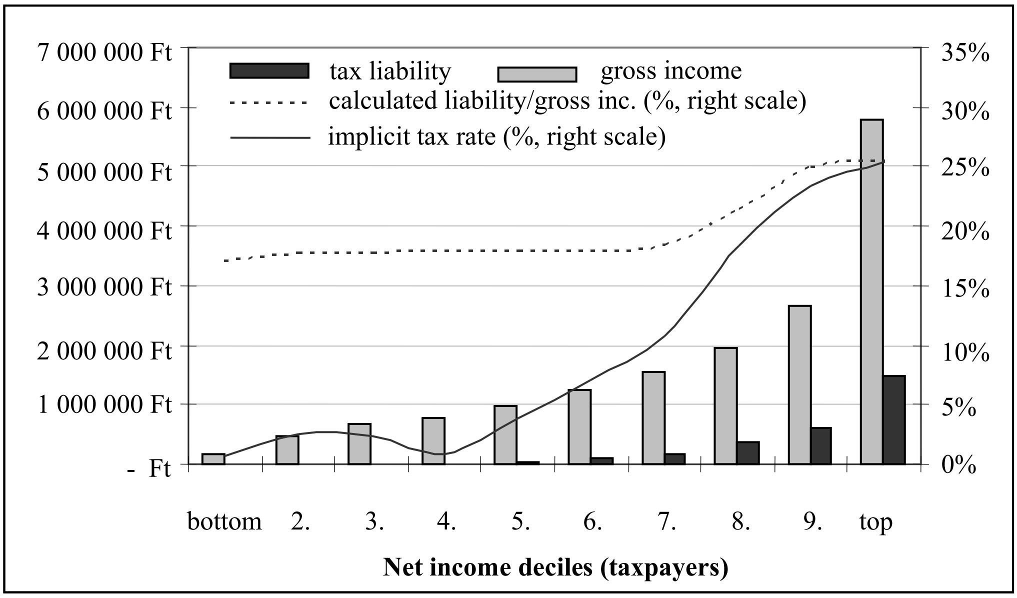


Figure 3. Gross income and tax liability, and the calculated implicit tax rate, 2006

Source: own calculations with TÁRSZIM2005

Note: implicit tax rate = actual tax liability/gross income (i.e. including tax allowances); calculated liability = tax calculated by the tax scheme + capital taxes (i.e. without tax allowances)

The Hungarian budget relies heavily on indirect taxes, and received about 63% higher amount from VAT than from PIT in 2006 [Ministry of Finance, (2007)], a typical phenomenon of economies with high tax evasion as the collection of the former is somewhat easier for the administration. Burden from both taxes falls on households (although note that some of the VAT is paid by corporations), but the progressivity is rather different. While PIT is progressive, VAT is actually regressive, in the sense that the share of VAT within incomes decreases as income increases (Figure 4). The reason of this phenomenon is that rich households differ from poor ones in terms of their savings rather than their spending. Redistribution via VAT is further complicated by the fact that goods under the preferential VAT rate (e.g. basic food items) are consumed in a similar amount by the rich and poor households, therefore the preferential rate subsidises all households to a similar extent. The progressivity of the

VAT scheme might be also influenced by unit non-response in the consumption survey, as discussed before.

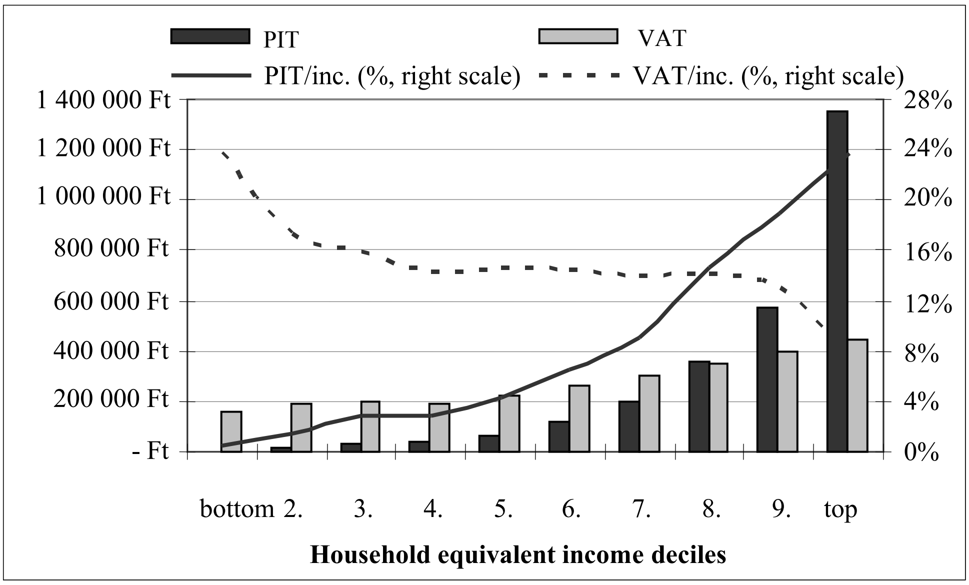


Figure 4. PIT and indirect tax (VAT and excise duties)
liability of households in forints and ratio of the disposable income

Source: own calculations with TÁRSZIM2005

The Hungarian data correspond to the theoretical propositions. In one of the first articles on this topic, Atkinson and Stiglitz (1976) argued that in an optimum income tax regime[ix] – if certain conditions are met with regard to the preferences of consumers – there is no need for indirect taxes (which may reduce the efficiency of the system). However it has been shown by several authors [e.g. Saez (2000)] that the information asymmetry between government and individuals still necessitates the application of indirect taxes. Another argument for indirect taxes is poor tax compliance, which precludes an optimum income tax regime. In the case of poor taxpayer morale, it is cheaper and easier to collect indirect taxes. Boadway, Marchand and Pestieau (1994) come to the conclusion that an optimum tax regime comprises a combination of direct and indirect taxes. With indirect taxes, however, for the sake of fairness, higher rates should be imposed on goods consumed more often by higher-income groups of society.

Although tax allowances seem to reach the poorest among the taxpayers, the picture is rather different when we focus on the disadvantaged within the whole population. The reason is rather intuitive: only a few households have labour income in the bottom decile, thus they cannot (fully) deduct tax allowances. In addition, pensioners, the majority of whom belong to the 4th to 7th deciles, are not entitled to most of these benefits either. On the other hand cash benefits do reach the poorest. It should be noted however that not all forms of tax allowances have poverty alleviation purposes, and some of them are not meant to be targeted to the poor. Figure 5 shows that the bottom decile receives a substantial amount of cash benefits, but cannot exercise tax allowances due to their low incomes. The second decile however can use both forms of support already. Nevertheless main beneficiaries of the tax allowances are in the top third of the distribution. It may not be surprising though, as tax allowances include preferential treatment of repayment of subsidised mortgage loans, life and pension savings, adult education, etc., which are often associated with the more affluent groups.

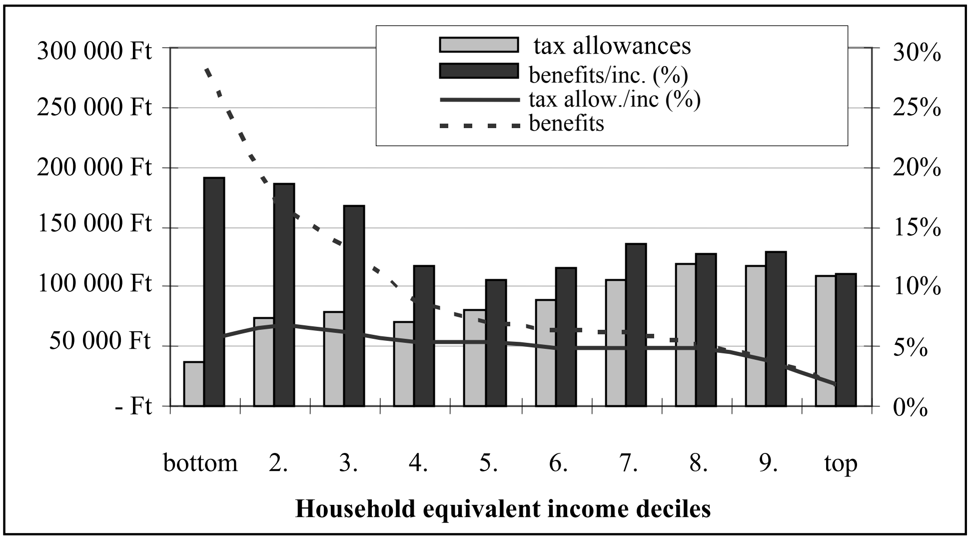


Figure 5 : Tax allowances and benefits per households in the deciles

Source: own calculations with TÁRSZIM2005

Employee tax allowance and family tax allowance are supposed to serve social purposes, but they do not reach the poorest decile sufficiently (Figure 6) either. The employee tax allowance is to guarantee that the minimum wage is exempt from PIT, and it is tapered away at higher income levels (the maximum amount of the allowance decreases to zero between 1.35 and 2 million HUF). Families with 3 or more children are eligible for family tax allowance. High proportion of these families belong to the lower deciles, therefore the targeting of this allowance is better, as a lower proportion goes to the upper half of the distribution, however the bottom 10 percent receives a lower proportion of this type of benefit than the second decile.

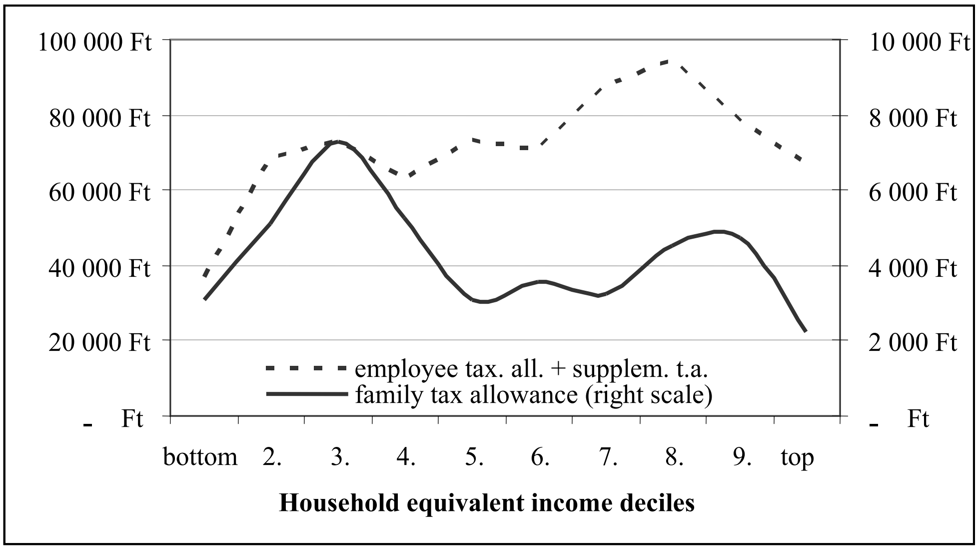


Figure 6: Average value of employee tax allowance and family tax allowance per households in various income deciles

Source: own calculations with TÁRSZIM2005

Note: the so-called "supplementary tax allowance" is added to employee tax allowance as they serve a similar purpose

The number of children is one of the main risk factors of poverty in Hungary. The universal child benefit, with its progressive amounts[x], is therefore rather successful in reaching the poorest income groups. In addition, this cash benefit is better targeted than tax allowances, which often cannot be fully deducted by those with low incomes. For example families with 4 or more children in the bottom decile group get about 50% of their total disposable income from this source, but even among the poorest 2-child families child support counts for about one third of their income. Although child support succeeds in reaching poor households well, the proportion of this benefit going to the upper half of the distribution is also remarkable, simply due to its universal nature.

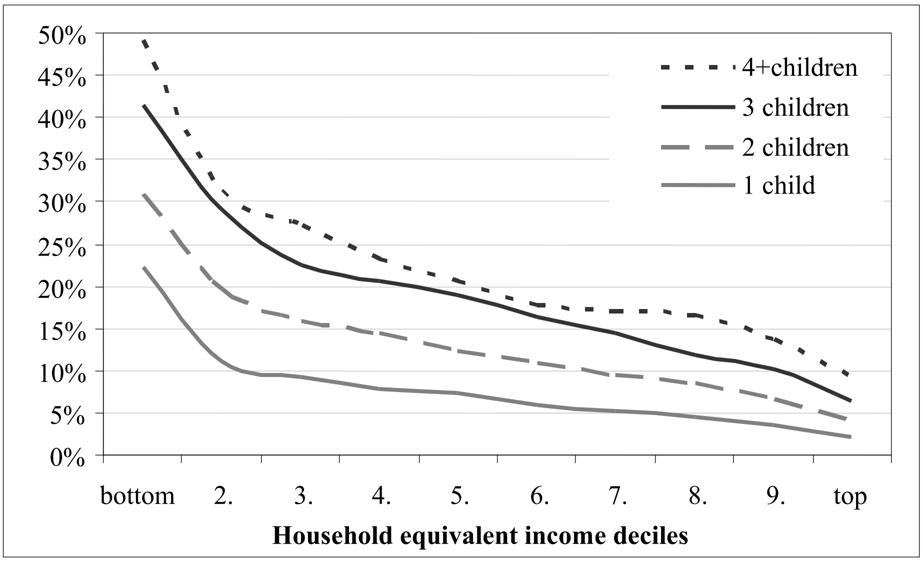


Figure 7: Proportion of the universal child benefit within total household income by income deciles and by number of children

Source: own calculations with TÁRSZIM2005

5. Effects of a flat tax system on incomes: gainers and losers

Flat tax appears to be a popular policy idea in Eastern Europe. The majority of Eastern European countries seem to have introduced such a scheme, starting with the Baltic states in the mid 1990s, followed by Russia, Ukraine, Georgia, Kazakhstan, Kyrgyzstan, Slovakia, Serbia, Romania, Macedonia later. Albania is planning to do so from 2008. Poland, and Greece is also considering such a move. In many of the countries it meant a simplification of the scheme, e.g. with the elimination of tax allowances, but also the lowering of average tax burden, although not in all of them. Lithuania, for example, set the flat rate at the highest bracket of the existing regime, 33%.

What is actually meant by "flat tax", varies a great deal per country. Very few countries have a flat tax scheme, where all types of incomes are taxed equally, including earnings, capital income and corporate income, and even less do tax consumptions to the same extent. Perhaps most strikingly, however, it would impose flat rates on people, irrespective of their income levels (and their abilities to conceal incomes, e.g. via transfers to foreign countries), thus giving up the role of the state in redistributing incomes via the tax system. This may save administrative costs, as all redistribution is costly, as claimed by Okun (1975) in his picturesque image of the leaky bucket. None of these countries, however, go for the "pure" version of it. In most cases, there are tax free limits for people with low incomes, at times also for families. Flat tax reform may have its appeal as it serves as an opportunity for simplifying the tax scheme, primarily from numerous tax allowances accumulated over long years due to interest groups rent-seeking. Simplicity saves takes administration costs and increases compliance.

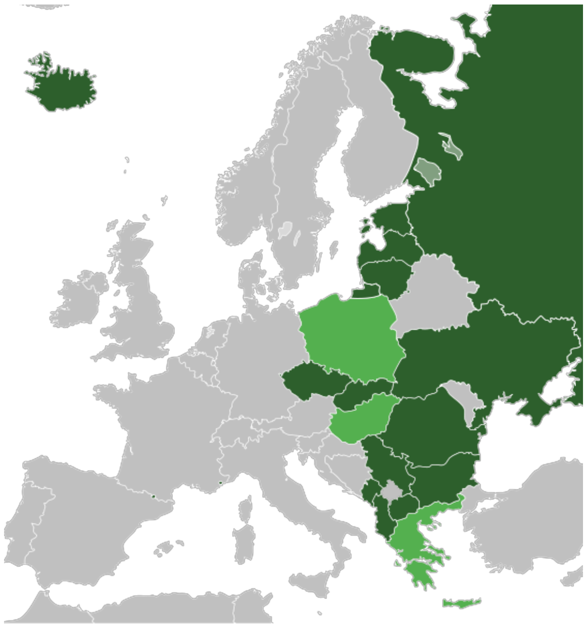


Figure 8: Flat tax systems in Europe, indicating countries which adopted a flat tax regime and those considering its introduction (Hungary, Poland, Greece)

Source: http://en.wikipedia.org/wiki/Flat_tax, Retrieved on 10 August 2008

Albeit the international popularity of the scheme, little is yet known on its impact on macro performance and on tax compliance. One of the success stories is Russia, where tax revenues have increased by 25% following the 2001 introduction of the scheme. Ivanova, Keen and Klemm (2005) find that the reform has probably increased tax compliance considerably (by one third) in Russia, it is not clear to what extent it is attributable to the parametric reform and to the increased law enforcement. The authors also highlight that oil revenues have increased massively in this period, resulting fast economic growth, thus producing greater tax revenues. Slovakia has introduced flat taxes as part of an overall structural reform, including pensions, social transfers, education, health care and the tax system. They introduced a uniform 19% tax rate (replacing 18 different rates), widened the tax base, and decreased, but not eliminated the progressivity of the tax system. The reception of the reform was positive by the international financial market, reflected in the improved country credit rating. On the other hand, there is no empirical evidence on the behavioral impact of the reform.

5.1 Microsimulation results

Referring to Hungary, we conducted a 'what if' type of analysis, where the tax regime parameters are entirely hypothetical. We take a tax regime that is somewhat similar to the one introduced in the course of tax reform in Slovakia, simpler than the existing Hungarian system: all tax rates (PIT and VAT) are set at 20%, there are no special tax rates but all incomes are combined and the flat rate is applied, the only allowance remaining is the employer tax allowance, the gas price subsidy is removed, but the means tested benefits, family benefits and insurance based benefits (including e.g. pensions) remain. The budgetary effects are not explored, yet we mean to present an example that is, on the whole, likely to be revenue neutral for the budget.

As mentioned before, the model is a static one, so it does not take account of behavioural responses. The economic outcomes are likely to be altered by behavioural responses. The virtue of such analysis is not prediction of long term outcomes, rather to show the day after effect, which is then likely to lead to certain actions. The policy relevance of such analysis is different, but equally relevant.

A 20% single tax rate would benefit primarily the higher-income groups. With the exception of the top fifth of the income distribution, there are no major differences in the PIT tax liability, the taxes paid by the richest would, however, be reduced substantially (Figure 9). VAT would increase in all households and the gainers of PIT and VAT changes together would be the rich while losers would be the poor.

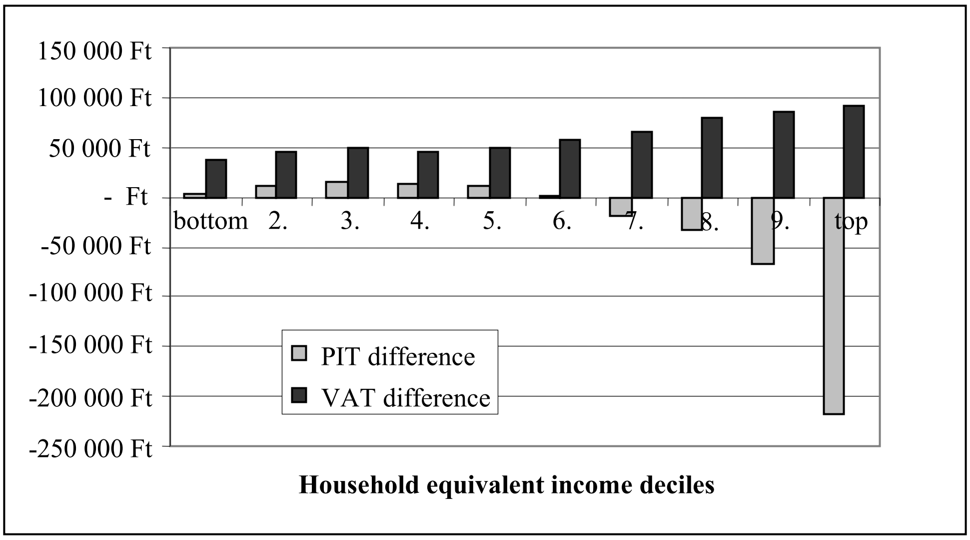


Figure 9: Increase in PIT and VAT liabilities of households compared to the 2006 system (2006 vs 20% flat tax) by income deciles

Source: own calculations with TÁRSZIM2005

In terms of disposable income, the poorest would be affected worst by such a tax simplification (Figure 10). On the whole, only the top third of the distribution would benefit significantly, while people in the bottom half would be left with smaller disposable incomes.

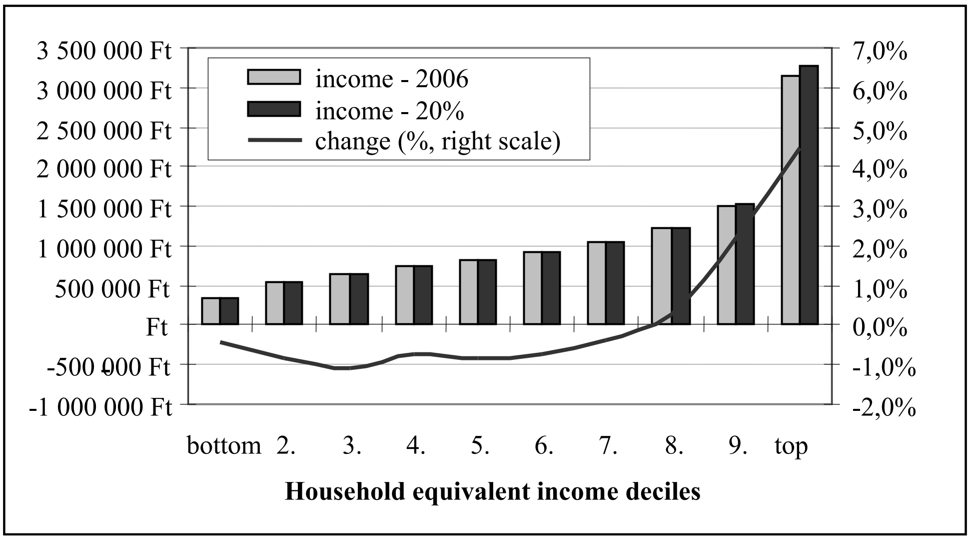


Figure 10: Equivalent disposable incomes in 2006 and after the tax simplification and % change

Source: own calculations with TÁRSZIM2005

Such a simplified tax system would benefit about 15% of households, while 15% of households would be worse off than at present. Income would not change substantially for the remaining households (Figure 11). Households gaining would be found at the top deciles while the proportion of losers is above 10% in each decile, but is the highest at the middle of the distribution. Therefore, the introduction of such a tax regime would reduce the extent of income redistribution.

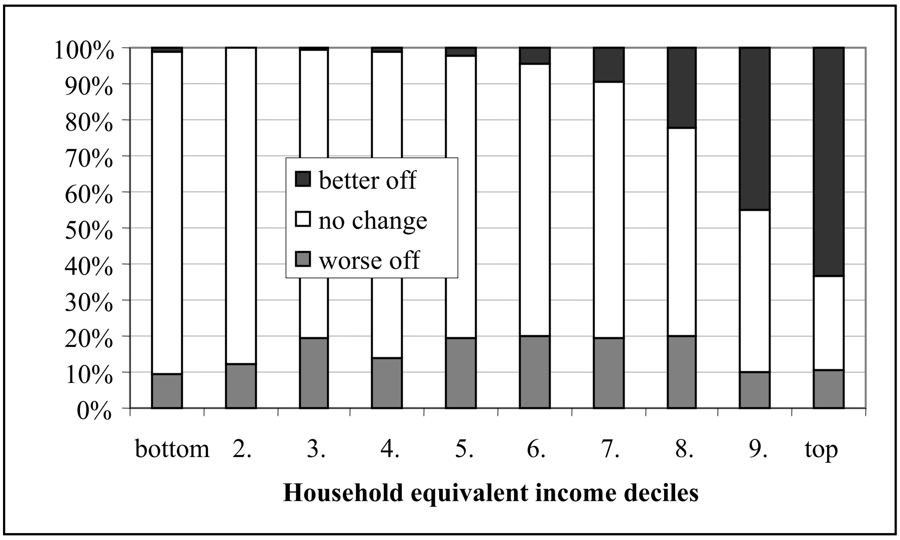


Figure 1. Distribution of households gaining/losing as a result of the changes

Source: own calculations with TÁRSZIM2005
Notes:
Worse off: whose per capita disposable income has dropped by more than 2%;
No change: whose per capita disposable income has not changed by more than 2%;
Better off: whose per capita disposable income has grown by more than 2%.

The political discussion and the debate of the theme should therefore be specific about the priorities and the parameters of such a reform, and any decision should be preceded by impact analyses of different kinds.

6. Conclusions

The paper presented evidence on the redistributive impacts of the current tax and benefit system and the possible effects of the introduction of a hypothetical 20% flat tax system in Hungary. There is little evidence on the size and impact of tax allowances, and they do not constitute part of the annual budgetary process. We showed that amount of tax allowances is substantial, and this type of benefits does not reach the poorest but rather the middle of the distribution. Cash benefits have a greater equalising role as they reach the poorest third of the population and significantly increase the disposable income of these groups. According to our calculations a Slovakian type flat tax system would mostly benefit the affluent households, while poorer losers are likely to lose out. The main reason is the changes in the income tax rates, with an increase at the lower and a decrease at the higher end, but the increase of the preferential VAT rate also plays a role. Future research options will significantly broaden by the incorporation of Hungary into the EUROMOD the European tax-benefit model [on Euromod see: Immervol, O'Donoghue and Sutherland, (1999); on the feasibility of enlargement Euromod to Hungary, see: Varga, Gabor, Szivos and Vajda (2006)]. Hungary is one of the four countries selected for the enlargement of the model as part of an ongoing project. The new version of the Euromod model would enable cross-country comparisons, including East and West, old and new member states. An interesting novel research area could be the estimation of the impact of tax evasion on income distribution. Microsimulation results, where the actual degree of tax compliance is taken into account rather than assuming full compliance, would provide more accurate insights into the restributive system as a whole.

7. Acknowledgements:

The authors would like to thank István Tóth for useful dicsussions and Prof John Hills and Prof Holly Sutherland for personal meetings on the subject of tax-benefit microsimulation as such.

8. References:

[1] Alm, J., (1996), *What Is an „Optimal Tax System?"*, in: *National Tax Journal,* Vol. 49 no. 1, 117 – 133.

[2] Atkinson, A.B., Stiglitz, J.E., (1976), *The design of tax structure: direct versus indirect taxation*, in: *Journal of Public Economics* 6, 55– 75.

[3] Benedek, D., Firle, R., and Scharle, Á., (2006), *Degree and efficiency of redistribution*, in: *Ministry of Finance Working Papers* No. 17, Hungary.

[4] Benedek, D., Lelkes, O., Scharle, Á. and Szabó, M., (2006), *A magyar államháztartási bevételek és kiadások szerkezete 1991-2004 között* [The structure of general government expenditure and revenues in Hungary, 1991-2004], in: *Közgazdasági Szemle* **53**: 119 – 143.

[5] Benedek, D., Lelkes, O., (2005), *Assesment of Income Distribution in Hungary using a Microsimulation model*, in: *Ministry of Finance Working Papers* No. 10, Hungary.

[6] Boadway, R., Marchand, M., and Pestieau, P., (1994), *Towards a Theory of Direct-Indirect Tax Mix*, in: *Journal of Public Economics* 55, 71 – 88.

[7] Förster, M.F., Tóth I. Gy., (1997), *Poverty, inequalities and social policies in the Visegrád countries*, in: *Economics of Transition* **5**(2): 505 – 510.

[8] Förster, M.F., and Tóth, I.G., (2001), *Child poverty and family transfers in the Czech Republic, Hungary and Poland*, in: *Journal of European Social Policy* **11**(4): 324 – 341.

[9]Gábos, N.A., Szivós, P., and Györgyi, V., (2006), *Euromod I-CUE Feasibility Study: Hungary. The 2005 Tax-Benefit System.* Vienna, European Centre, Retrieved on 10 July 2007 from http://www.euro.centre.org/icue

[10] Immervol, H., O'Donoghue, C., Sutherland, H., (1999), *An introduction to EUROMOD*, in: *EUROMOD Working Paper* No. EM0/99.

[11] Ivanova, A., Keen, M., and Klemm, A., (2005), *The Russian 'Flat Tax' Reform*, in: *Economic Policy* 20(43): 397 – 444.

[12] Central Statistical Office, (2004), *Háztartás-statisztikai közlemények*, 25. szám, 2004. I. félév, Central Statistical Office, Budapest.

[13] Lelkes, O., (2000), *A great leap towards liberalism? The Hungarian welfare state*, in: *International Journal of Social Welfare* **9**(2): 92 – 102.

[14] Lelkes, O., (2006), *Tasting Freedom: Happiness, religion and economic transition*, in: *Journal of Economic Behavior and Organization* **59**(2): 173 – 194.

[15] Lelkes, O., (2007), *Tax-benefit microsimulation models in Eastern Europe*, in: *International Journal of Microsimulation.* 1(1): 54 – 56.

[16] Ministry of Finance, Hungary, (2007), *A központi költségvetés előzetes mérlege."* [Balance sheet of the central government, preliminary figures]. Retrieved on 15 July 2007 from www.pm.gov.hu

[17] Okun, A.M., (1975), *Equality and Efficiency: the Big Trade-Off*, Washington, The Brookings Institution.

[18] Redmond, G., (1999), *Incomes, incentives and the growth of means-testing in Hungary*, in: *Fiscal Studies* **20**(1): 77 – 99.

[19] Saez, E., (2002), *The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes*, in: *Journal of Public Economics* 83, 217 – 230.

[20] Sutherland, H., Levy, H., and Paulus, A., (2006), *The effects of taxes and benefits on income distribution in the EU*, in: *Final Report of the Network on Social Inclusion and Income Distribution*, European Observatory on the Social Situation, financed by the European

Commission (DG Employment, Social Affairs and Equal Opportunities) http://ec.europa.eu/employment_social/social_situation/docs/sso2005_social_inclusion_report.pdf

[21] Szivós P., Rudas, T., Tóth, I. Gy., (1998), *TÁRSZIM97 Mikroszimulációs modell az adók és támogatások hatásvizsgálatára*, [Microsimulation model for analysing the impact of taxes and benefits]. *TÁRKI Társadalompolitikai tanulmányok* 10., TÁRKI, Budapest.

[22] TÁRSZIM, (2005), Professional microsimulation modell v3.2. Manual. TÁRKI Rt., Budapest.

[23] Tóth, I. Gy., (2005), *Jövedelemeloszlás. A gazdasági rendszerváltástól az uniós csatlakozásig,* [Income Distribution. From Economic Transition to the Accession to the European Union], Budapest, Századvég Kiadó. Andorka Rudolf Társadalomtudományi Társaság.

9. Notes:

i. In Hungary, benefits make up 10% of gross incomes among the top quintile, while this ratio is about 20% in Austria and France.

ii. TÁRSZIM2005 model is the product of TÁRKI Social Research Inc., Budapest – the software was developed by VirgoSystems Kft. -, financed by the Ministry of Finance and Ministry of Social Affairs, using the datasets of APEH, CSO and TÁRKI.

iii. On the potential applications of EUROMOD see the publication series at http://www.iser.essex.ac.uk/msu/emod/publications/emodwp.php.

iv. For an overview of microsimulation models in Eastern Europe, see Lelkes (2007).

v. On previous microsimulation modelling in Hungary see Szivós-Rudas-Tóth (1998), description of earlier model version of TÁRSZIM is presented on the relevant web page of TÁRKI: http://www.tarki.hu/research/mikro/index.html. See also Redmond (1999) on the results of the common model of CSO and University of Cambridge.

vi. Including the consumption of health care into someone's cash incomes would imply that people who are sicker would be shown as richer, simply due to the consumption of these services, with no obvious consequences of this „richness" on their well-being or utility.

vii. A detailed technical description of the compilation of the data file can be found in the Handbook of TÁRSZIM2005 Professional v3.2.

viii. For a detailed description of the Hungarian tax and benefit system see Benedek, Firle and Scharle (2006), and the structure of government taxes and spending, see Benedek, Lelkes, Scharle and Szabo 2006.

ix. An optimum tax regime is characterised by the following: as a result of market failures (external effects) the state needs to intervene in market mechanisms by applying taxes, i.e. the state needs to raise tax revenue. An optimum arrangement among the different tax systems is one that entails the smallest distortion or social cost (including primarily 'dead weight loss'). For more details on the subject see: James Alm (1996).

x. The amount of benefit *per children* actually increases by the number of children (up to three children, where it stays the same for additional children). For example, a family with three children receives 3.8 times higher benefit than a one-child family.

DOES VOLATILITY IN GOVERNMENT BORROWING LEADS TO HIGHER INFLATION? EVIDENCE FROM PAKISTAN

S. Adnan, H.A.S. **BUKHARI**[1] and Safdar Ullah **KHAN**[2]
Research Department, State Bank of Pakistan, **PAKISTAN**
adnan.haider@sbp.org.pk[1] and safdar.khan@sbp.org.pk[2]

Abstract:

This study analyzes the impact of volatility in government borrowing from central bank (GBCB) on domestic inflation in Pakistan. This paper utilizes Generalized Auto Regressive Conditional Hetroskedasticity (GARCH) model to estimate volatility in GBCB using monthly data from July 1992 to June 2007. The empirical results, based on auto regressive distributed lag (ARDL) with bound testing technique suggest that domestic inflation in Pakistan is related with volatility in government borrowing from central bank in the long run. Furthermore, error correction model (ECM) estimates show that in the short run, inflation is also affected by volatility in GBCB.

Keywords: inflation, government borrowing, volatility, GARCH, ARDL, ECM

JEL Classifications: E31, E60, E62

1. Introduction

> *"A common criticism of this stress on the budget deficit is that the data rarely shows a strong positive association between the size of the budget deficit and the inflation rate."*
>
> (Blanchard and Fischer, 1989, p. 513)

Historical literature recognizes inflation as a monetary vis-à-vis fiscal phenomenon with momentous socio-economic and political consequences. Generally, inflation refers to a *sustained increase in general price level* as measured by an index such as consumer price index (CPI) or gross domestic product (GDP) deflator. It may be either demand-pull or cost-push.[2]

Economists identify inflation with different categories depending on the degree of variability in inflation. For instance, when the general price level is rising by 5% or less, they call it creeping inflation. This creeping inflation provides health to the economy by generating the forces of dynamism. If the prices become stagnant, the economic growth will stop in the economy. When the general price level is rising by a rate between 5 to 10%, it is labeled as a situation of walking inflation. This walking inflation gives a signal that something is wrong in the economic management process and proper remedial measures have to be initiated. If the increase in general price level is above 10%, the situation is called running inflation which becomes hyper inflation when it shooting more than 50%.

Presumably, inflation generates welfare cost of economic agents (i.e., inflation tax) and inflation volatility is considered as a key source of destabilizing mistakes. It frequently varies and thus increases uncertainties for macroeconomic environment. Rother (2004) argue that high variability of inflation over time makes expectations over the future price level more uncertain. In a world with nominal contracts this induces risk premia for long-term arrangements, raises costs for hedging against

---

[1] Corresponding author: Research Department, State Bank of Pakistan, I. I. Chundrigar Road, Karachi, Pakistan. Views expressed here are those of the authors and do not necessarily of the State Bank of Pakistan. Author welcomes comments and suggestions.

[2] Demand-Pull Inflation is situation often described as *too much money chasing too few goods*. According to this view, an excess of aggregate demand over aggregate supply will generate inflationary pressures in prices. Cost-Push Inflation is caused by wages increase by union pressures and profit increase by producers. The basic cause of cost-push inflation is that the money wages increase more rapidly than the productivity of labor. Cost-push inflation may be due to upward adjustment of wages to compensate rise in the cost of living index. An increase in the prices of domestically produced or imported raw materials may lead to cost-push inflation. Another cause of cost-push inflation is increase in easy and non-functional profits by oligopolistic and monopolist firms.

inflation risks and leads to unanticipated redistribution of wealth. To some prominent economists inflation is always and everywhere a monetary phenomenon in the long run. That is why almost all over the world central banks are entrusted upon to tame inflation. For all the central banks taking care of inflation is at least one of the objectives of monetary policy. Same is the case with the State Bank of Pakistan (SBP) being central bank of the country. SBP Act, 1956 confers upon it to regulate money and credit in the country in such a way that maintains monetary stability (which leads to price stability) while fostering the utilization of country's resources in the best national interest.

A general consensus among macroeconomist is that inflation occurs when the rate of growth of the money supply is higher than the growth rate of the economy.[3] This phenomenon, however, occurs usually in developing countries which faces high budget deficits. The central government of any developing country finances their budget deficit through monetizing process (borrowing from central bank). High monetization leads to higher inflationary pressure to the economy. Thus borrowing from the central bank is considered as a leading indicator of domestic inflation.

In line with the above phenomenon, the main motivation of this paper is to assess whether volatility in government borrowing has an impact of domestic inflation in Pakistan. For this purpose Generalized Auto Regressive Conditional Hetroskedasticity (GARCH) model is used to estimate volatility in government borrowing from central bank (GBCB) using monthly data from July 1992 to June 2007. Our main hypothesis is that GBCB has significant and positive impact on domestic inflation. It is also one of the leading indicators of price instability both in short and in the long run.

The rest of the paper is organized as follows: A review of previous empirical studies is presented in section 2. Section 3 provides data and methodology. Empirical findings are discussed in section 4 and the main conclusions are stated in section 5.

2. Brief Review of Empirical Literature

There is an immense literature available on fiscal vis-à-vis monetary determinants of inflation. In this paper we also provide a brief review of some selected domestic and international studies. This review provides us stylized facts and a baseline for our model consistency. Therefore, we intend to categorize the literature for Pakistan into two sets including studies which used government borrowing as a determinant of inflation and those which have not incorporated this determinant in their model setup. Appendix A summarizes almost all of the local empirical literature on inflation since 1982 to date. In this study we also present evidence of international empirical literature relevant to this concept.

In the case of Turkey, Akcay, Alper and Ozmucur (1996) investigate determinants of inflation using annual data from 1948 to 1994 vis-à-vis quarterly data from 1987 to 95. Their analysis reveals that a one unit increase in the deficit GNP ratio under money neutrality will increase the long-run inflation by 1.59 units. Also a one unit increase in the deficit GNP ratio under money neutrality will increase the long-run inflation by 5.67 which is much higher than 1.59 for the whole sample indicating greater impact of deficit on inflation during pre-bond financing period.

Metin (1998) provide a multivariate cointegration analysis of the determinants of inflation for Turkey using annual data from 1950 to 1987. The major finding from the new equation is that budget deficits (as well as real income growth and debt monetization) significantly affect inflation. For the conditional model, an increase in the scaled budget deficit immediately increases inflation. Real income growth has a negative immediate effect and positive second-lag effect on inflation. Monetization of the deficit also affects inflation at a second lag.

Catao and Terrones (2003) studied the deficit-inflation relationship in 107 countries over the period 1960 to 2001. This study was distinctive in two respects. Firstly, it used an intertemporal optimization model to show that the equilibrium inflation is directly related to fiscal deficit which is scaled by narrow money. This approach resulted in introducing nonlinearity in the model which is better than semi-logarithm specification used earlier4. Secondly, they modeled the link between fiscal deficit and inflation "as intrinsically dynamic, explicitly distinguishing between the short run and long

[3] This is the conventional monetarist linkage from the creation of reserve money to inflation when Central Banks issues money at the rate that exceeds the demand for cash balances at the existing price level and the increased demand in the goods market pushes up the price level as the public tries to get rid of its excess cash holdings.

[4] This non-linearity in the model resulted in capturing a stronger impact of fiscal deficit in higher inflation economies.

run". This study specified an autoregressive distributed lag (ARDL) model for each country and polled them together in a panel, and then tested the cross-equation restriction of a common long-run relationship between the two variables using pooled mean group estimator (PMG)5. As the difference between the MG and PMG estimates of long-run elasticity parameter identified sample heterogeneity, therefore the authors divided the panel into groups on the basis of financial development and inflation performance. Then estimates of MG and PMG indicated that budget deficit was significant driver of inflation in most groups except the low inflation economies and advanced economies.

The results showed that in case of developing countries, a reduction (increase) of one percent in ratio of budget to GDP lowered (raised) inflation by around 8.75 percentage points. For emerging market economies, a percentage point change in the ratio of budget balance to GDP is estimated to change inflation by 2.25 percentage points. Similarly, changes in inflation strongly impacted the high inflation economies, and less strongly on moderate inflation economies. This study concluded that fiscal deficit displayed a powerful effect on inflation in developing countries, emerging markets and high-inflation economies and a much smaller effect amongst moderate inflation countries.

Rother (2004) examined the relationship between discretionary fiscal policies and inflation volatility for fifteen industrialized countries for the period from 1967 to 2001. Their results suggested that the volatility in discretionary fiscal policies strongly contributed to inflation volatility. They found that a one standard deviation increase in discretionary fiscal policy could raise inflation volatility to range of 10-17 percentage points. These results were obtained using panel data and performing regressions for different measures of inflation volatility (conditional and unconditional variability of inflation rate6) as a function of the volatility of activist fiscal policies7 and other explanatory variables (like output gap, monetary and exchange rates). Moreover, the Generalized Least Squares (GLS) was employed in order to account for the possibility of cross-sectional Hetroskedasticity.

Alavirad and Athawale (2005) investigate determinants of inflation in Islamic Republic of Iran using annual data from 1963 to 1999. They find that budget deficits do have a significant impact on inflation rates in the long run in the Islamic Republic of Iran. The ECM results show that the budget deficit and liquidity in the short run, and related to the long run, have less of an effect on price levels. The coefficient of error correction is estimated at -0.2. This value shows that the adjustment speeds is relatively slow.

Catao and Terrones (2005), using panel of 23 emerging market countries for the period 1970-2000 to investigate determinants of inflation. They found that a one percentage point reduction in the ratio of fiscal deficit to GDP lowered long-run inflation by 1.5 to 6 percentage points. This study used an econometric specification derived from inter-temporal optimization model that relates to long-run inflation to the permanent component of fiscal deficit. One of the most distinguishing feature of this study is that it fiscal deficit is scaled by the size of inflation tax base which is measured by the ratio of narrow money to GDP. This resulted in introduced the desired non-linearity in the relationship between fiscal deficit and inflation.

Giannitsarou and Scott (2006) examined the means through which fiscal sustainability was achieved by six industrialized countries (namely US, Japan, Germany, Italy, UK and Canada) during the period 1960-2005. They assessed the relative contribution of primary deficit, inflation and GDP growth as means to counter fiscal imbalances in the countries under study. Their findings suggested that fiscal balance was achieved mainly through variations in primary deficits (80-100 percent), whereas inflation (0-10 percent) and GDP growth (0-20 percent) contributed minimally towards fiscal sustainability.

The empirical results suggested that fiscal imbalances and weak forecaster for future inflation in economies under study. More specifically, they found that the predicted rise in fiscal deficit scenario

[5] This methodology is better than earlier used static fixed-effects estimator as country-specific ARDL structure is capable of accommodating cross-country heterogeneity in inflation inertia.

[6] Unconditional variability in inflation rate is defined as the standard deviation over a calendar year of month-on-month inflation rates, thereby capturing the extent of short-term fluctuations in inflation. Moreover, conditional inflation variability is measured by the standard deviation of one-step-ahead forecast errors derived from time-series based inflation forecast model.

[7] To measure the volatility of discretionary fiscal policy, fiscal policy stance is defined as the year-on-year change in the cyclically adjusted primary balance relative to GDP.

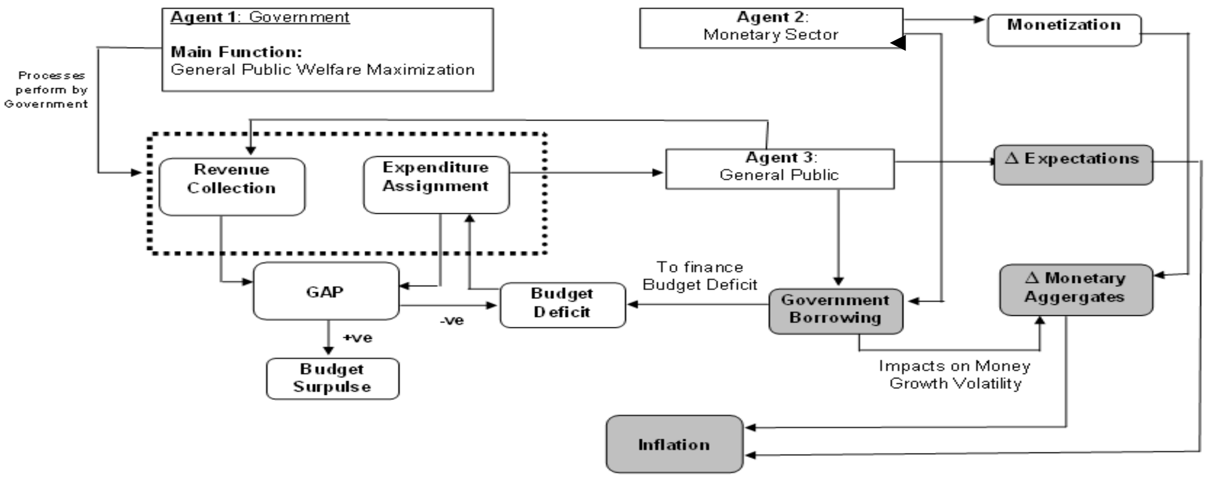
in future could possibly impact in an insignificant manner towards increasing inflation in the economy. The authors further observed that their results should be used with much caution as econometrically evaluating the inter-temporal budget constraint is vulnerable to non-stationarity and time dependence problems.

3. Methodology and Data Description

3.1 Theoretical Framework

To study the adverse impact of volatility in government borrowing from central bank on domestic inflation, it is necessary to observe its functional channel (see, flow chart 1). Theoretically speaking, budget deficit (*BD*) weakly causes inflationary pressures, but rather impacts strongly on general price level through the impact on money aggregates (say, M1 and M2) and public expectations, which in turn trigger volatility in prices. Since, government borrows from different sources to finance budget deficits, so it is necessary to observe its dynamics which generate volatility in money growth. To do this, we adopted a theoretical model introduced by Sachs and Larrain (1993).

Flow Chart 1. Functional Channel of Government Borrowing and Money Growth Volatility



The budget constraint of the public sector government as introduce by Sachs and Larrain (1993) can be expressed as follows:

$$BDF = \Delta GD^{g} = GD^{g} - GD^{g}_{-1} = P\cdot(G + I^{g} - T) + i\cdot GD^{g}_{-1} \tag{3.1}$$

Where:

- $GD^{g} - GD^{g}_{-1}$ *, is the change in government debt between the current and previous periods,*
- *P is the price level,*
- $G + I^{g}$ *, is Government expenditures*
- *T is taxes*
- $i\cdot GD^{g}_{-1}$, is the interest payments on previously issued debt.

Government debt, in the form of either bonds or credits, can be held by the public (domestic and foreign) and by the central bank. Let's assume for the purposes of the present report that the central bank's credit to banking system doesn't alter over time. Then the change in monetary base ΔMB equals the change in the stock of government debt held by central bank $(GD^{g}_{c} - GD^{g}_{c-1})$ plus the change in foreign exchange reserves, $E\cdot(B^{*}_{c} - B^{*}_{c-1})$, where E stands for the nominal exchange rate, we obtain:

$$\Delta GD^g = \Delta MB + (GD_p^g - GD_{p-1}^g) - E \cdot (B_c^* - B_{c-1}^*) \qquad (3.2)$$

Equation (3.2) gives us information that there are three ways to cover a budget deficit; [a]: by "monetization" of the deficit (i.e. by increasing monetary base or by so called "printing" money); [b]: by increase in the public's (foreign and domestic) holdings of debt; and [c]: by running down foreign exchange reserves at the central bank. Since, our target is to find the volatility in government borrowing from central bank on domestic inflation. So for simplicity in our model, we assumed that government only borrows from central bank[8]. In this case, equation (3.2) becomes:

$$\Delta GD^g = \Delta MB \qquad (3.3)$$

Where;

$$(GD_c^g - GD_{c-1}^g) = 0 \quad \text{and} \quad E \cdot (B_c^* - B_{c-1}^*) = 0$$

Let:

$$\Delta GD^g = GB_g \quad \text{and} \quad \Delta MB = M_g$$

Where:

- GBg is growth in government borrowing;
- Mg is the money growth

This type of borrowing is called "monetizing" [9] the deficit. Because this phenomenon always leads to the growth of monetary base (MB) and money supply, it is often defined as "printing money". From equation (3.3), we can observe that an increase in the high-powered money is the source of financing budget deficit.

Lastly, from equation (3) we can define general functional form as:

$$Mg = f(GB)$$

Or

$$volt(Mg) = volt(GB_g) \qquad (3.4)$$

Equation (3.4) implies that volatility in government borrowing impacts directly on money growth via monetization channel. Quantity theory also identify that the volatility in money growth is the key factor that effect the changes in price level [Walsh (2003) and Romer (2006)]. So, we also have the following relationship:

$$\pi_t = volt(MB_g) \qquad (3.5)$$

Where; π_t → Domestic inflation

Hence, from relationship (4) and (5) we can also establish a direct relationship that volatility in government borrowing have an impact on domestic inflation as:

---

[8] Ouanes and Thakur (1997) argues that there exist five different ways of financing budget deficit: (a) borrowing from the central bank (or "monetization" of the deficit); (b) borrowing from the rest of the banking system; (c) borrowing from the domestic non-bank sector; (d) borrowing from abroad, or running down foreign exchange reserves; and (e) accumulation of arrears.

9 Monetization occurs (i) when the central bank directly finances budget deficit by lending funds needed to pay government bills; or (ii) when the central bank purchases government debt at the time of issuance or later in the course of open market operations.

$$\pi_t = volt(GB_g) \qquad (3.6)$$

In order to estimate the functional relationship (3.6), we use ARCH/GARCH model introduced by Engle (1982) and Bollerslev (1986), respectively. To apply this methodology, it is necessary to search for appropriate type of ARCH/GARCH specifications to model the dynamics of GBCB volatility. We apply LM test developed by Engle [1982] to determine time varying volatility behavior as well as searching for the asymmetric effects of shocks on volatility.

Consider an information set Ω about GBCB. So, jointly estimated standard ARCH/GARCH model is given as:

$$GB_t = \mu + \varepsilon_t \qquad (3.7)$$

Where; $\varepsilon_t = \sigma_t z_t$ and $z_t \sim iid(0,1)$

$$\sigma_t^2 = \omega + \psi\varepsilon_{t-1}^2 + \phi\sigma_{t-1}^2$$

Using model specifications (3.6) and (3.7), we have a final version of our complete econometric model.

$$\left.\begin{aligned} \pi_t &= \alpha + \beta vol(GB_t) + \xi_t \\ GB_t &= \mu + \varepsilon_t \\ \sigma_t^2 &= \omega + \psi\varepsilon_{t-1}^2 + \phi\sigma_{t-1}^2 \end{aligned}\right\} \qquad (3.8)$$

Where; $\varepsilon_t = \sigma_t z_t$ and $\xi_t, z_t \sim iid(0,1)$

3.2 Data Description

This paper uses Auto Regressive Distributed Lag (ARDL) test with bound testing technique to investigate the long run relationship between volatility in GBCB[10] and domestic inflation using Pakistan's time series data taken from the *Pakistan Economic Survey* and *Annual Reports* (various issues) and SBP monthly statistical bulletins (various issues) . This data series is on monthly basis from 1992 to 2007. In line with our hypothesis, we also provide some stylized facts in Figure 1 to Figure 3 shows a positive correlation between government borrowing from central bank and domestic inflation.

[10] Generalized Auto Regressive Conditional Hetroskedasticity (GARCH) model is used to estimate volatility in GBCB.

Fig 1: Government Borrowing from SBP and Inflation

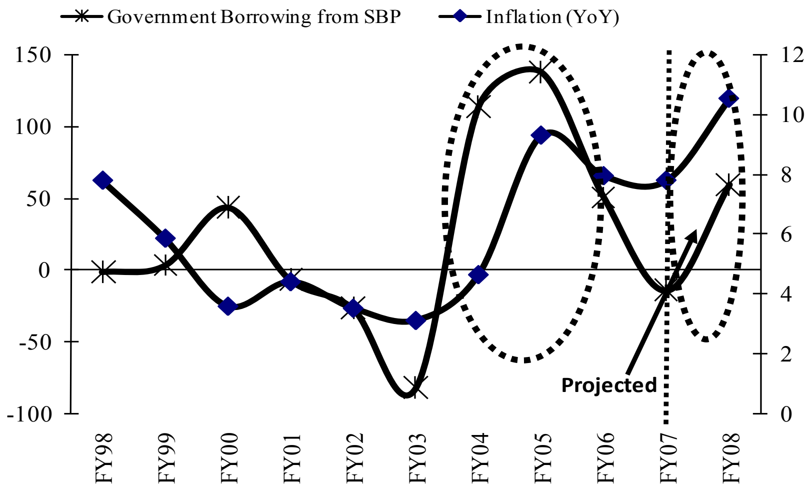


Fig 2: Budget Deficit Growth and Inflation

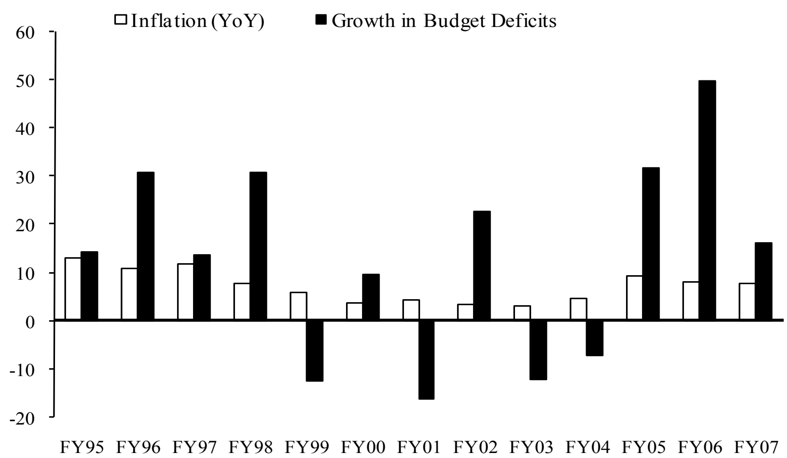
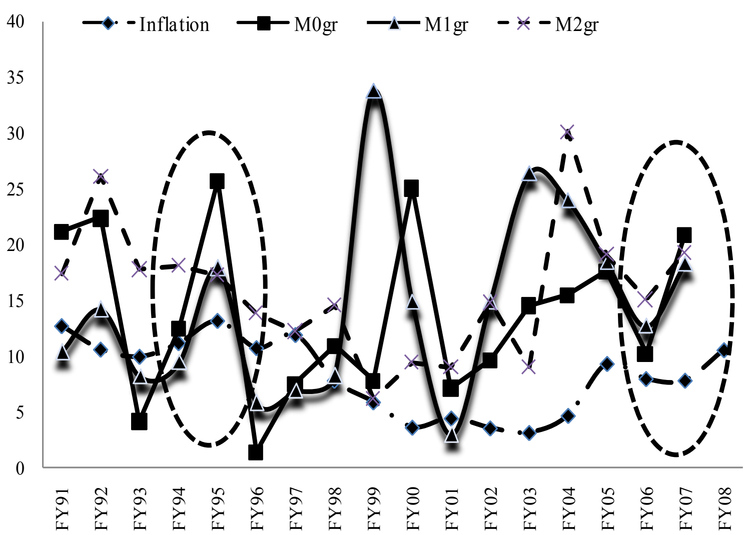


Fig 3: Inflation and Monetary Aggergates



Note: M0gr := Growth in Reserve Money; M1gr := Growth in M1 and M2gr := Growth in M2

Furthermore, the process of monetization is also be observed from growth in monetary aggregates which also leads domestic inflation in Pakistan. The whole graphical representation clearly provide us a sketch that government finance its budget deficit with borrowing from central bank which lead pressure on monetary aggregates and hence rises domestic inflation in Pakistan. This also shows that recent higher government borrowing from central bank leads higher domestic inflation in Pakistan, since FY07.

4. Empirical Results and Discussion

In line with our study hypothesis, we estimate model (3.8) using monthly data from July 1992 to June 2007. Estimation results of model (3.8) are presented in Table 1. We use log difference form of consumer price index as a proxy of domestic inflation for Pakistan. The results in Table 1 provide us important information that the impact of volatility in government borrowing from central bank on domestic inflation is economically and statistically significant. The estimated coefficient of government borrowing shows that one standard deviation change in (volatility) government borrowing from central bank leads 8.5% change (increase) in domestic inflation Pakistan.

Now in order to investigate long run dynamics (cointegration) between domestic inflation and volatility in government borrowing, we use ARDL model as introduce by Pesaran et al (1999). Detail methodological description is also available in Appendix B. In line with this methodology we are going to introduce cointegration functional form as:

$$\Delta\pi_t = \alpha_0 + \sum_{i=1}^{k}\alpha_i\pi_{t-i} + \sum_{i=0}^{k}\beta_i volt(GB)_{t-i} + \sum_{i=1}^{k}\gamma_i\Delta\pi_{t-i} + \sum_{i=0}^{k}\delta_i\Delta volt(GB)_{t-i} \quad (3.9)$$

Two further aspects of the regression equation (3.9) need specifying in practice. First we specify the lag order 'k' in the regression. We started testing with a maximum lag of 12 and used information criteria and sequential *F tests* along with tests for residual autocorrelation to guide our lag choice. Since this is monthly data and we wish to preserve as many degrees of freedom as possible, this seems a reasonable maximum lag order. The second decision regards the inclusion of deterministic constant and trend terms. We report here tests based on a model with an unrestricted constant, since we found no evidence of a significant deterministic trend in the relationship. We based our decision on lag order on the observation of information criteria, F test of the reduction (from 12 lags to 1 lag) and the autocorrelation test. Tests of the null hypothesis of no long run relationship can thus be carried out using an F test of the null that

$\alpha_i = \beta_i = 0$.

Results in Table 2 suggest there is a strong long run relationship between domestic inflation and volatility in government borrowing from central bank. The value of F-statistic shows a significance of the rejection of null hypothesis of no cointegration as suggested in Pesaran, Shin and Smith (1999)[11].

Finally, since the above result appears to confirm the existence of a long run relationship, we use the estimated regression to form an error correction term and estimate a simple dynamic ECM for domestic inflation. The estimated regression is reproduced below with a standard range of diagnostics. The results of ECM model are presented in Table 3.

The error correction term is correctly signed and significant. The value of the coefficient on the ECM indicates that a change in volatility in government borrowing from central bank brings about a 77% change in domestic inflation in Pakistan in the span of twelve months. The ECM also passes a range of diagnostic tests.

[11]This method is, once again, applicable irrespective of whether the regressors are I(0) or I(1). The long run estimates and their standard errors were obtained using EViews 5.0.

Table 1. ARCH/GARCH Model Estimation Results
Impact of volatility in GBCB on Domestic inflation

Parameters	μ	ω	ψ	ϕ
Estimated Coefficients	-1.031	96.984	2.651	-0.013
S.E	0.610	9.321	0.183	0.008
t-ratios	-1.982	10.505	14.508	-1.998
$\pi_t = 0.518 + 0.085 vol(GB_t)$ (7.232) (1.994)				

*Note: Values in prentices shows t-statistics

Table 2. F test for the existence of a long run relationship

Test Statistic	Value	df	Probability
F-statistic	40.150	(4, 137)	0.0000
Chi-square	80.301	4	0.0000

*Note: 95% critical bounds for the F test: 4.94 - 5.73[12]

Table 3. Error correction model Results

Variable	Coefficient	S.E	t-value
Constant	-0.005	0.051	-0.113
$\Delta volt(GB)_t$	0.018	0.034	0.532
$\Delta \Pi_{t-1}$	0.161	0.094	1.686
$\Delta volt(GB)_{t-1}$	0.023	0.034	0.664
ECM_{t-1}	-0.767	0.085	-9.849

$R^2 = 0.4208$, D.W = 2.02

5. Conclusion

This study supports the fiscal dominance hypothesis in determining inflation in the case of Pakistan economy. In connection with this hypothesis, the results depict important information that the impact of volatility in government borrowing from central bank on domestic inflation is economically and statistically significant. Further, the empirical evidence suggests that there is a strong long run relationship between domestic inflation and volatility in government borrowing from central bank. The estimated coefficient of government borrowing shows that one standard deviation change in (volatility) government borrowing from central bank leads 8.5% change (increase) in domestic inflation. In particular, it suggests incorporating the trend effects of government borrowing from monetary authorities in inflation modeling. Finally these findings may help in understanding inflation experience in different developing economies like Pakistan.

6. References

[1] Agha, A.I., and Khan, M.S., (2006), *An Empirical analysis of Fiscal Imbalances and inflation in Pakistan*, in: *SBP research Bulletin*, Vol. 2, No. 2, pp. 343 – 362.

[2] Ahmad, E., and Harim, R., (1991), *Foreign Price Shocks and Inflation in Pakistan: A Monetarist Approach*, in: *Pakistan Economic and Social Review,* Vol. 29, No. 1, pp. 1–20.

[3] Ahmad, E., and Munirs, Muhammad, (2000), *An Analysis of Money Demand in. Pakistan,* in: *Pakistan Economic and Social Review*, Vol. 38, No. 1, pp. 47–67.

[4] Ahmad, E., and Saima, A.A., (1999a), *Exchange Rate and Inflation Dynamics,* in: *The Pakistan Development Review*, Vol. 28, No. 3, pp. 235–51.

[5] Ahmad, E., and Saima A.A., (1999b), *Relationship between Exchange Rate and Inflation*, in: *Pakistan Economic and Social Review*, Vol. 37, No. 2, pp. 139–54.

[12] Critical bounds are from Table C1.iii of Pesaran *et al*., (1999).

[6] Ahmad, Mushtaq, and Khan, Ashfaque H., (1990), *A Reexamination of the Stability of the Demand for Money* in: *Pakistan*, in *Journal of Macroeconomics*, Vol. 12, No. 2, pp. 307–21.

[7] Akcay, O.C., Alper, C.E., and Ozmucur, S., (1996), *Budget Deficit, Money Supply and Inflation: Evidence from Low and High Frequency Data from Turkey*, Bogazici University Research Papers, SBE 96 – 12.

[8] Akinboade, E.W., Niedermeier, E.W, and Siebrits, F.K., (2003), *South Africa's Inflation Dynamics: Implications for Policy*, Paper presented in: *75th Conference of the Economic Society of South Africa*, Glenburn Lodge.

[9] Akinboade, E.W., Siebrits, F.K., and Niedermeier, E.W., (2001), *The Determinants of Inflation in South Africa: An Econometric Analysis*, Final Research Report submitted to the African Economic Research Consortium, Nairobi, Kenya.

[10] Alavirad A. and Athawale, Sanhita, (2005), *The impact of the budget deficit on inflation in the Islamic Republic of Iran*, in: *OPEC Review*, Vol. 29 No. 1, pp. 37-49.

[11] Arize, A.C., (1994), *A Re-Examination of the Demand for Money in Small Developing Economies*, in: *Applied Economics*, Vol. 26, pp. 217–28.

[12] Aslam, C.M., and Naveed, A., (1995) *Money Supply, Deficit, and Inflation in Pakistan*, in: *Pakistan Development Review*, Vol. 34, No. 4, (III), pp. 945 – 956.

[13] Aslam, C.M., and Shahid, W.A., (1996), *Macroeconomic policies and Management of Debt, Deficit, and inflation*, in: *Pakistan*, in *Pakistan Development Review*, Vol. 35, No. 4, (II), pp. 773 – 786.

[14] Banerjee, A., and Bill, R., (2002), *Inflation and Measures of the Markup*, Manuscript, Department of Economics, European University Institute, The Department of Economic Studies, University of Dundee.

[15] Bailliu, G., Kruger, A., and Messmacher, E., (2003), *Explaining and Forecasting Inflation*, in: *Emerging Markets: The Case of Mexico*, working paper 2003 – 17, Banco de México.

[16] Bilquees, F., (1981), *Inflation in Pakistan: An Empirical Analysis*, PIDE Research Report no 119, Pakistan Institute of Development Economics, Islamabad, Pakistan.

[17] Bilquees, F., (1988), *Inflation in Pakistan: Empirical Evidence on the Monetarist and Structuralist Hypotheses*, in: *The Pakistan Development Review*, Vol. 27, No. 2, pp. 109–130.

[18] Blejer, M., (1977), *The short run dynamics of prices and the balance of payments*, in: *American Economic Review*, Vol. 67, No. 3, pp. 419 – 428.

[19] Bokil and Schimmelpfennig, (2005), *Three Attempts at Inflation Forecasting*, in: *Pakistan*, IMF Working Paper No. WP/05/105.

[20] Bollerslev, T., (1986), *Generalized conditional Hetroskedasticity*, in: *Journal of Econometrics*, Vol. 31, pp. 307 – 327.

[21] Bowdler C. and Jansen. E.S., (2004), *Markup Model of Inflation for the Euro Area*, ECB Working Paper No 36/February/2004.

[22] Burney, N.A., and Akmal, M., (1990), *Expected Inflation Rate, Its Variability, and Desired Real Money Demand in Pakistan: Some Alternative Estimates Using Non-Linear Squares Method*, in: *Pakistan Economic and Social Review*, Vol. 28, No. 1, pp. 69 – 88.

[23] Catao, L. and Terrones, M., (2001), *Fiscal Deficits and Inflation: A New Look at the Emerging Market Evidence*, IMF Working Paper, WP/01/74.

[24] Catao, L., and Marco, E.T., (2005), *Fiscal Deficits and Inflation*, in: *Journal of Monetary Economics*, Vol. 52, No. 2, pp. 529 – 54.

[25] Choudhri, E.U., and. Khan, M.S, (2002), *The Exchange Rate and Consumer Prices in Pakistan: Is Rupee Devalution Inflationary?,* in: *Pakistan Development Review*, Vol. 41 No. 1, pp. 107 – 120.

[26] Dhakal, D., and Kandil, M., (1993), *The Inflationary Experience of Six Developing Countries in Asia: An Investigation of Underlying Determinants*, in: *Applied Economics*, Vol. 25, No. 2, pp. 413 – 25.

[27] Engle, R.F., (1982), *Auto Regressive conditional Hetroskedasticity with estimates of the variance of United Kingdom inflation*, in: *Econometrica*, Vol. 50, pp. 987 – 1007.

[28] Giannitsarou, C., and Scott, A., (2006), *Inflation Implications of Rising Government Debt*, NBER Working Paper No. 12654.

[29] Gunnar, J., (2001), *Inflation, Money Demand, and Purchasing Power Parity in South Africa*, IMF Staff Papers, Vol. 48, No. 2, International Monetary Fund.

[30] Hamburger, M.J., and Zwick, B., (1982), *Deficit, Money and Inflation, Reply,* in: *Journal of Monetary Economics*, Vol. 10, No. 2.

[31] Hasan, M.A., Khan, A.H., Pasha, H.A., and Rashid, M.A., (1995), *What Explains the Current High Rate of Inflation in Pakistan*?, in: *Pakistan Development Review*, Vol. 34, pp. 927 – 943.

[32] Hossain, A., (1986), *Monetary Disequilibrium and Inflation: a monetary Model of Inflation in Pakistan*, in: *Pakistan Development Review*, Vol. 25, No. 2, pp. 141 – 162.

[33] Hossain, A., (1994), *The Search for a Stable Money Demand Function for Pakistan: An Application of the Method of Cointegration*, in: *The Pakistan Development Review* Vol. 33 no. 4.

[34] Hsing, Yu, (1998), *Tests of Changes in the Elasticity of the Demand for M2 and Policy Implications: The Case of Four Asian Countries*, in: *Journal of Economic Development*, Vol. 23, No. 2, pp. 181 – 89.

[35] Huq, M.D., Shamsul and Majumdar, Badiul A., (1986), *Stability of the Demand for Money Function: Evidence from a Developing Country*, in: *Pakistan Economic and Social Review*, Vol. 24, No. 1, pp. 45 – 55.

[36] Junggun, A., (2000), *Inflation Targeting: A New Monetary Policy Framework in Korea*, Paper prepared for International Symposium on Practical Experiences on Inflation Targeting, Bangkok, Thailand.

[37] Khan, A.A., Bukhari, S.K.H., and Q. M. Ahmad (2006), *Determinants of Recent inflation in Pakistan*, SPDC Research Report.

[38] Khan, Ashfaque, (1982a), *An Analysis of the Demand for Money in Pakistan*, in *Journal of the Institute of Bankers in Pakistan*, Vol. 48, pp. 25 – 34.

[39] Khan, Ashfaque H., (1982b), *The Demand for Money and the Variability of the Rate of Inflation - An Empirical Note*, in *Economics Letters*, Vol. 10, pp. 257 – 61.

[40] Khan, A.H. (1994), *Financial liberalization and the demand for money in Pakistan*, in: *Pakistan Development Review,* Vol. 33, No. 3, pp. 997 – 1006.

[41] Khan, A.H., and Qasim, M.A., (1996), *Inflation in Pakistan Revisited*, in: *The Pakistan Development Review*, Vol. 35, No. 4, pp. 747 – 759.

[42] Khan, Imran Naveed, (1992), *The Demand for Money in Pakistan & India*, in: *Pakistan Economic and Social Review*, Vol. 30, No. 2, pp. 181 – 90.

[43] Khan, M.S., and Schimmepfennig, A., (2006), *Inflation in Pakistan: Money or Wheat?*, IMF Working Paper, WP/06/60.

[44] Metin, K., (1998), *The Relationship Between Inflation and Budget Deficit in Turkey*, in *Journal of Business & Economic Statistics*, Vol. 16, No.4, pp. 412 – 421.

[45] Naqvi, Syed Nawab, Kemal, A.R., and Rashid, Aziz, (1982), *The P.I.D.E. Econometric Model of Pakistan's Economy (1959-60 to 1978-79)*, Published by Pakistan Institute of Development Economics, Islamabad, Pakistan.

[46] Naqvi, S.N.H., and Khan, A.H., (1989), *Recent Inflationary Upsurge in Pakistan*, in *Pakistan Development Review*, vol. 28, No. 4, pp. 762 – 791

[47] Naqvi, S.N.H., Khan, A.H., Ahmed, A.M. and Siddiqui, R., (1994), *Inflation in Pakistan: Causes and Remedies*, Research Report, Pakistan Institute of Development Economics, Islamabad.

[48] Nasim, A. (1997), *Determinants of Inflation in Pakistan*, SBP Publication, State Bank of Pakistan Press, Karachi, Pakistan.

[49] Nisar, Shaheena, and Naheed, Aslam, (1983), *The Demand for Money and the Term Structure of the Interest Rates in Pakistan*, in *The Pakistan Development Review*, Vol. 22, No. 2, pp. 97 – 116.

[50] Ouanes, A., and Thakur, S., (1997), *Macroeconomic Accounting and Analysis in Transition Economies*, Washington, D.C.: International Monetary Fund.

[51] Pesaran, M.H., and Shin, Y., (1995), *Long-run Structural Modeling*, DAE Working Paper No. 9419, Department of Applied Economics, University of Cambridge, Cambridge.

[52] Pesaran, M.H., Shin, Y., and Smith, R.J., (2001), *Bounds Testing Approaches to the Analysis of Level Relationships*, in *Journal of Applied Econometrics*, Vol. 16, pp. 289 – 326.

[53] Pesaran, M.H., and Pesaran, B., (1997), *Working with Microfit 4.0: An Interactive Approach*, Oxford: Oxford University Press.

[54] Price, S., and Nasim, A., (1999), *Modelling Inflation and the Demand for Money in Pakistan; Cointegration and the Causal Structure*. In *Economic Modelling*, Vol. 16, pp. 87 – 103.

[55] Qayyum, A., and Bilquees, F., (2005), *P-Star Model: A Leading Indicator of Inflation for Pakistan*, in *The Pakistan Development Review*, Vol. 44, No. 2, pp. 117 – 129.

[56] Rother, P.C., (2004), *Fiscal Policy and Inflation Volatility*, ECB Working Paper Series, WP No. 317/March/2004.

[57] Romer, D., (2006), *Advanced Macroeconomics*. Third Edition, New York: The McGraw-Hill Companies, Inc.

[58] Sachs, J.D., and Larrain, F., (1993), *Macroeconomics in the Global Economy*, Englewood Cliffs, NJ: Prentice-Hall, Inc.

[59] Shamsuddin, A.F.M., and Holmes, R.A., (1997), *Cointegration Test of the Monetary Theory of Inflation and Forecasting Accuracy of the Univariate and Vector ARMA Models of Inflation*. In *Journal of Economic Studies*, Vol. 24, pp. 294 – 306.

[60] Siddiqui, R., and Akhtar, N., (1999), *The Impact of Changes in Exchange Rate on Prices: A Case Study of Pakistan*, in *Pakistan Development Review*, Vol. 38, No. 4.

[61] Tariq, S.M., and Kent Matthews, (1997), *The Demand for Simple Sum and Divisia Monetary Aggregates for Pakistan: A Cointegration Approach*, in *Pakistan Development Review*, Vol. 36, No. 3.

[62] Walsh, E.C., (2003), *Monetary Theory and Policy*, Second Edition, MIT Press, Cambridge, England.

Appendix A

Table A1- Selected Pakistan Empirical Studies of Inflation and Monetary Policy

Authors	Empirical Approach	Dependent Variable(s)	Regressors	Sample Period	Findings
Section (i): Pakistan Studies which used Government Borrowing as a determinant of Inflation					
Agha, Asif Idrees and Khan, Muhammad Saleem (2006)	Johansen cointegration analysis, VECM model	consumer price index	consolidated fiscal deficit, total bank borrowing	1973 to 2003	The empirical results suggest that in the long-run inflation is not only related to fiscal imbalances but also to the sources of financing fiscal deficit.
Khan, A. Aleem, Bukhari, S. K. Hyder and Ahmed, Q. Masood (2006)	Ordinary least square (OLS) method and verifying results through Breusch-Godfrey Serial Correlation LM and Augmented Dickey-Fuller tests	consumer price index	Governmnet sector borrowing (plus NFA and other items) as ratio to real GNP, real demand relative to real supply, non-government sector borrowing (pluc borrowing of autonomous bodies) as ratio of real GNP, price index of imports, exchange rate, government taxes as a ratio of manufacturing sector value added, lagged CPI and support price of wheat.	1972 to 2005	The most important determinants of inflation are adaptive expectations, private sector credit and rising import prices whereas fiscal policy's contribution to inflation was minimal. Specifically, if government sector borrowing as a ratio to GNP chnaged by 10 percent, then the resulting change in CPI will be around 1 percent.
Chaudhary, M. Aslam and Anjum, S. Waseem (1996)	Sustainable deficit econometric model for Pakistan is estimated.	Growth rate of GNP, inflation rate, interest rate of foreign debt etc.	A number of assumptions regarding growth rate of GNP, inflation rate and interest rate on foreign debt	Three time periods-1980s, 1985-95 and 1993-98	Throughout the period under analysis, fiscal defict was not sustainable.
Chaudhary, M. Aslam and Ahmed, Naved (1995)	Simultaneous model and OLS i.e regressions of money supply equation, real cash balance equation, price equation, output equation and export supply equation.	Consumer Price index, money supply, demand for real cash balances, exports	(1) Money supply equation- international reserves, domestic financing of budget deficit including banking and non-banking system, commercial banks credit to private sector (2) Demand for real cash balances- income, proxy for cost of holding real balances (3) Price equation- income, money supply, import price (4) Output equation- governmnet expenditures, commercial credit (5) Export supply equation- income, export price	1973-92, 1973-82 and 1982-92	Domestic financing of budget deficit, particularly from the banking system is inflationary in long run. Money supply is not exogenous, rather it depends on the position of international reserves and fiscal deficit and it has emerged as an endogenous variable.
Section (ii): Pakistan Studies which did not consider Government Borrowing as a determinant of Inflation					
Authors	**Empirical Approach**	**Dependent Variable(s)**	**Regressors**	**Sample Period**	**Findings**
Hyder, Zulfiqar and Sardar Shah (2004)	VAR		CPI inflation, WPI inflation, PR/USD, M2, LSM index, oil prices	1988:1 to 2003:9	Little exchange rate pass through to domestic CPI inflation.

Choudhri, Ehsan U. and Mohsin S. Khan (2002)	Single equation and VAR in first differences	CPI and WPI	U.S. dollar exchange rate, foreign price index	1982–2001	There is no exchange rate pass-through to domestic prices.
Ahmad, Eatzaz and Muhammad Munirs (2000)	OLS, cointegration analysis	M1, M2	Index of industrial production, interbank call money rate, CPI inflation	1972:I to 1996:I	Find that inflation is a better measure of opportunity cost than interest rate, money demand adjusts sluggish, and there was a structural break in the early 1990s.
Ahmad, Eatzaz and Saima Ahmed Ali (1999a)	Single equation, including Engle/Granger cointegration test, 2-equation model with 2SLS	CPI and exchange rate	Exchange rate, import prices, world prices, money supply, GDP, forex reserves	1982:II to 1996:IV	CPI reacts to changes in import prices (due to change in world prices or exchange rate) and money supply. Exchange rate responds to domestic and world prices.
Ahmad, Eatzaz and Saima Ahmed Ali (1999b)	2-equation model with 2SLS	CPI and exchange rate	Exchange rate, import prices, world prices, money supply, GDP, forex reserves	1982:II to 1996:IV	CPI reacts to changes in import prices (due to change in world prices or exchange rate) and money supply. Exchange rate responds to domestic and world prices.
Price, Simon and Anjum Nasim (1999)	Johansen (VECM), and SUR	CPI and exchange rate	Broad money, world prices, GDP, deposit rate	1974 to 1994	PPP and money demand relation are identified that are connected through cointegrating relationships.
Hsing, Yu (1998)	Single equation	Real M2	Real GDP, deposit rate	n.a.	Real GDP elasticity is close to unity whereas interest elasticity is low.
Shamsuddin, Abul F.M. and Richard A. Holmes (1997)	Johansen procedure, VARMA, ARMA	CPI	Broad money, real output	1972:II to 1993:IV	Rejects a cointegrating relationship between inflation, broad money and GDP and concludes that a univariate ARMA yields the best forecasts.
Tariq, Syed Muhammed and Kent Matthews (1997)	Johansen, single equation ECM	M2, M1, divisia	Real GDP, opportunity costs	1974:IV to 1992:IV	Identifies a cointegration vector that is interpreted as a money demand function. Short-run parameters of money demand equation are stable.
Chaudhary, M. Aslam and Naved Ahmad (1996)	OLS	CPI inflation	Broad money, GDP growth, share of service sector, public debt, import prices	1972 to 1992	Inflation results from money growth and structural factors such as growth, share of service sector, public debt, and import prices.
Arize, A.C. (1994)	OLS of ECM	M1, M2	GDP, inflation rate, call money rate, government bond yield, expected rate of depreciation (foreign interest differential)	1973:I to 1990:I	Finds that money demand is a function of GDP, inflation, interest rate and exchange rate expectations. Also, dummies for the oil shocks, and structural free banking and floating the rupee matter. account of introduction of partial interest- breaks in 1981 on
Hossain, Akhtar (1994)	Engle/Granger 2-stage, Johansen	M1, M2	GDP, yield on government bonds, market call rate, CPI inflation	1951–91	Meaningful cointegration relationship (money demand function) for the post- 1972 period.

Khan, Ashfaque H. (1994)	Engle/Granger 2-stage	M1, M2	Real income, real interest rate (short-term and medium-term), nominal interest rate (short-term and medium-term), inflation	1971:III to 1993:III	Finds cointegrating relationship between M2 (or M1) and real income, real interest rate and inflation.
Dhakal, Dharmendra and Magda Kandil (1993)	OLS of distributed lag specification (AIC)	CPI inflation	M1, industrial production, interest rate, foreign interest rate, import prices	1970:I to 1987:IV	Import prices, industrial production, and U.K. interest rate explain inflation. M1 is insignificant.
Khan, Imran Naveed (1992)	OLS	M1, M2	GNP, call rate, CPI	n.a.	Money demand in Pakistan is a function of income and inflation, but not of interest rate.
Ahmad, Eatzaz and Harim Ram (1991)	OLS	WPI, CPI, GNP deflator, and absorption deflator inflation	Real GNP growth, growth rate of unit value of imports, growth rate of M1/M2, lagged inflation	1960 to 1988	Inflation is determined by real GNP growth, unit value of import growth, nominal money growth, and lagged inflation.
Ahmad, Mushtaq and Ashfaque H. Khan (1990)	ML (Cooley/Prescott 1976 varying parameter technique)	M1, M2	Income, inter-bank call rate, time deposit rate	1959 to 1987	Demand for real money was unstable at th time of delinking the Pakistani rupee from the U.S. dollar and introduction of interest- free deposit accounts.
Burney, Nadeem A. and Mohammad Akmal (1990)	NLLS	Real money stock	Income, CPI inflation, CPI inflation volatility	n.a.	Real money adjusts instantaneously to the desired level of money demand which is driven by income, and expected inflation.
Khan, Ashfaque H. and Bilquees Raza (1989)	OLS	M1, M2	Real GNP, interest rate, expected inflation	1972:II to 1987:II	Larger than unity income elasticities of money demand and the expected influence of expected inflation and interest rates.
Huq, M.D. Shamsul and Majumdar, Badiul A. (1986)	OLS	M1, M2	GNP, call money rate, government bond rate, CPI inflation	1955 to 1977	Structural breaks in the demand for money in 1965 and 1971.
Nisar, Shaheena and Naheed Aslam (1983)	OLS	M1, M2	GNP, term structure, GNP deflator	1959 to 1978	Term structure matters for money demand besides income.
Khan, Ashfaque (1982a)	OLS	M1, M2	GNP, interest rate on time deposits	1959/60 to 1979/80	Income elasticity of 1.7 and interest elasticity of 0.5.
Khan, Ashfaque H. (1982b)	OLS	M1, M2	GNP, expected inflation, inflation variability	n.a.	Including the variability of inflation improves the estimate of the money demand function.
Naqvi, Syed Nawab, A.R. Kemal, and Rashid Aziz (1982)	53-equation macro model			1959/60 to 1978/79	Inflation is not imported. Money demand is interest-sensitive. The GNP elasticity of money demand is fairly large.

Appendix B

Auto Regressive Distributed Lag Model (ARDL MODEL)

Following Pesaran, Shin and Smith (1999) cointegration methodology using Auto Regressive Distributed Lag model this paper try to find long run relationship between inflation and volatility in government borrowing from central bank in Pakistan. This tests procedure is applicable irrespective of whether the regressors are I(0), I(1) or mutually cointegrated. The test is based upon estimation of the underlying VAR model, re-parameterised as an ECM(error correction model)[13].

The VAR(p) model $\mathbf{z}_t = \mathbf{b} + \mathbf{c}t + \sum_{i=1}^{p} \mathbf{\Phi}_i \mathbf{z}_{t-i} + \boldsymbol{\varepsilon}_t$ (A2.1)

where **z** represents a vector of variables. Under the assumption that the individual elements of **z** are at the most I(1), or do not have explosive roots, equation (A2.1) can be written as a simple Vector ECM.

$$\Delta\mathbf{z}_t = \mathbf{b} + \mathbf{c}t + \Pi\mathbf{z}_{t-1} + \sum_{i=1}^{p-1} \mathbf{\Gamma}_i \Delta\mathbf{z}_{t-i} + \boldsymbol{\varepsilon}_t \quad \text{(A2.2)}$$

where $\mathbf{\Pi} = -(\mathbf{I}_{k+1} - \sum_{i=1}^{p} \mathbf{\Phi}_i)$ and $\mathbf{\Gamma}_i = -\sum_{j=i+1}^{p} \mathbf{\Phi}_j$, i = 1,....p - 1 are the (k+1)×(k+1) matrices of the long run multipliers and the short run dynamic coefficients. By making the assumption that there is only one long run relationship amongst the variables, Pesaran *et al* focus on the first equation in (A2.2) and partition $\mathbf{z}_t$ into a dependant variable y_t and a set of forcing variables **x**. This is one of the key assumptions of their paper. Under such conditions the matrices **b, c Γ** and, most importantly, **Π,** the long run multiplier matrix can also be partitioned conformably with the partitioning of z.

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{11} & \boldsymbol{\pi}_{12} \\ \boldsymbol{\pi}_{21} & \mathbf{\Pi}_{22} \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ \mathbf{b_2} \end{bmatrix} \qquad \mathbf{c} = \begin{bmatrix} c_1 \\ \mathbf{c_2} \end{bmatrix} \quad \mathbf{\Gamma_i} = \begin{bmatrix} \gamma_{11,i} & \boldsymbol{\gamma}_{\mathbf{12,i}} \\ \boldsymbol{\gamma}_{\mathbf{21,i}} & \boldsymbol{\gamma}_{\mathbf{22,i}} \end{bmatrix}$$

The key assumption, that **x** is long run forcing for y, then implies that the vector $\gamma_{21}=0$, that is that there is no feedback from the level of y on Δ**x**. As a result the conditional model for Δy and Δ**x** can be written as

$$\Delta y_t = b_1 + c_1 t + \pi_{11} y_{t-1} + \pi_{12}\mathbf{x}_{t-1} + \sum_{i=1}^{p-1} \gamma_{11,i}\Delta y_{t-i} + \sum_{i=0}^{p-1} \boldsymbol{\gamma}_{12,i}\Delta\mathbf{x}_{t-i} + \varepsilon_{1t} \quad \text{(A2.3)}$$

$$\Delta\mathbf{x}_t = \mathbf{b_2} + \mathbf{c_2}t + \mathbf{\Pi}_{22}\mathbf{x}_{t-1} + \sum_{i=1}^{p-1} \gamma_{21,i}\Delta y_{t-i} + \sum_{i=1}^{p-1} \mathbf{\Gamma_{22,i}}\Delta\mathbf{x}_{t-i} + \boldsymbol{\varepsilon}_{2t} \quad \text{(A2.4)}$$

Under standard assumptions about the error terms in (A2.3) and (A2.4)[14] Pesaran *et al* re-write (A2.3) as

$$\Delta y_t = a_0 + a_1 t + \phi y_{t-1} + \delta\mathbf{x}_{t-1} + \sum_{i=1}^{p-1} \upsilon_i \Delta y_{t-i} + \sum_{i=0}^{p-1} \varphi_i \Delta\mathbf{x}_{t-i} + \varpi_t \quad \text{(A2.5)}$$

which they term an unrestricted error correction model. Note that in (A2.5) a long run relationship will exist amongst the levels variables if the two parameters ϕ and δ are both non zero in which case, for the long run solution of (A2.5) we obtain

$$y_t = -\frac{a_0}{\phi} - \frac{a_1}{\phi} - \frac{\delta}{\phi} x_t \quad \text{(A2.6)}$$

Pesaran *et al* choose to test the hypothesis of no long run relationship between y and x by testing the joint hypothesis that $\phi = \delta = 0$ in the context of equation (A2.5). The test they develop is a bounds type test, with a lower bound calculated on the basis that the variables in **x** are I(0) and an upper bound on the basis that they are I(1). Pesaran et al (1999) provide critical values for this bounds test from an extensive set of stochastic simulations under differing assumptions regarding the appropriate inclusion of deterministic variables in the ECM. If the calculated test statistic (which is a standard F test for testing the null that the coefficients on the lagged levels terms are jointly equal to zero) lies above the upper bound, the result is conclusive and implies that a long run relationship does exist between the variables. If the test statistic lies within the bounds, no conclusion can be drawn without knowledge of the time series properties of the variables. In this case, standard methods of testing would have to be applied. If the test statistic lies below the lower bound, no long run relationship exists.

[13] Most of the following is based on Pesaran, Shin and Smith (1999) and follows their original notation.

[14] Essentially that they are independently normally distributed with a positive definite variance covariance matrix.

THE DRIVING FORCE OF LABOR FORCE PARTICIPATION IN DEVELOPED COUNTRIES

Ivan O. **KITOV**
Institute for the Geospheres' Dynamics
Russian Academy of Sciences, **RUSSIA**
ikitov@mail.ru
Oleg I. **KITOV**
Warwick University, **UNITED KINDOM**

Abstract:

The evolution of labor force participation rate is modeled using a lagged linear function of real economic growth, as expressed by GDP per capita. For the U.S., our model predicts at a two-year horizon with RMSFE of 0.28% for the period between 1965 and 2007. Larger part of the deviation between predicted and measured LFP is explained by artificial dislocations in measured time series induced by major revisions to the CPS methodology in 1979 and 1989. Similar models have been developed for Japan, the UK, France, Italy, Canada, and Sweden.

Keywords: *labor force participation, real GDP per capita, prediction*

JEL Classification: C2, E6, J2

1. Introduction

The number of people having some paid job or its equivalent and those who are currently unemployed but seek for a job is called the labor force. The ratio of the labor force and overall working age population defines macroeconomic term “participation rate”, which provides a measure of labor supply not dependent on population size. Both working age population and labor force participation rate (LFP) vary over time and across countries. Obviously, the level of population of 16 years of age and older is driven by a multitude of factors including social, economic, geographic, and ethnic ones. Conventional economic theories also consider various forces driving the rate of labor force participation [Aaronson *et al*., (2006); Aaronson, Park, and Sullivan, (2006, 2007); Hausman, (1986); Haveman *et al*., (1991), Hotchkiss, (2004, 2005); Juhn and Potter, (2006); Murphy and Topel, (1987); Veracierto, (2008); Wachter, (1978); among many others]. These theories imply that the evolution and effects of these forces are difficult to forecast [Congressional Budget Office [CBO], (2004)]. Results of our research support an opposite view and demonstrate that there exists a unique factor completely controlling the evolution of participation rate in developed countries – real economic growth as defined by GDP per capita.

Because of the benefits provided by continuous, extensive, and open statistic information we first analyze and model the rate of labor force participation in the U.S.

Figure 1 displays the growth rate of the LFP obtained from its original time series [Bureau of Labor Statistics [BLS], (2008)] and a centered five-year moving average, MA(5). The rate rose during the 1960s through 1980s. Since 2000 it has been decreasing. Previous investigations, Aaronson *et al*. (2006) and Fullerton (2003) among others, demonstrated that, historically, the change of trend (for example, from the current negative one to a positive one) has not been accurately predicted.

Several models have been developed for the prediction of LFP at various time horizons. Aaronson *et al*. (2006) proposed a model with cyclical and structural components of the participation rate evolution affected by demographic factors, the business cycle, and other factors. Fallick and Pingle (2007) proposed a model of labor force participation based on cohorts. In their framework, the probability of individual participation declines with age beyond 50. However, Kitov (2005c) showed that there exists an effect of the increasing age of the peak mean income, which counteracts the effect of lowering participation with age. In any case, one can expect significant changes in the LFP due to the ageing of labor supply. Fallick and Pingle also considered the effects of participation trends on the aggregate LFP within all other age groups.

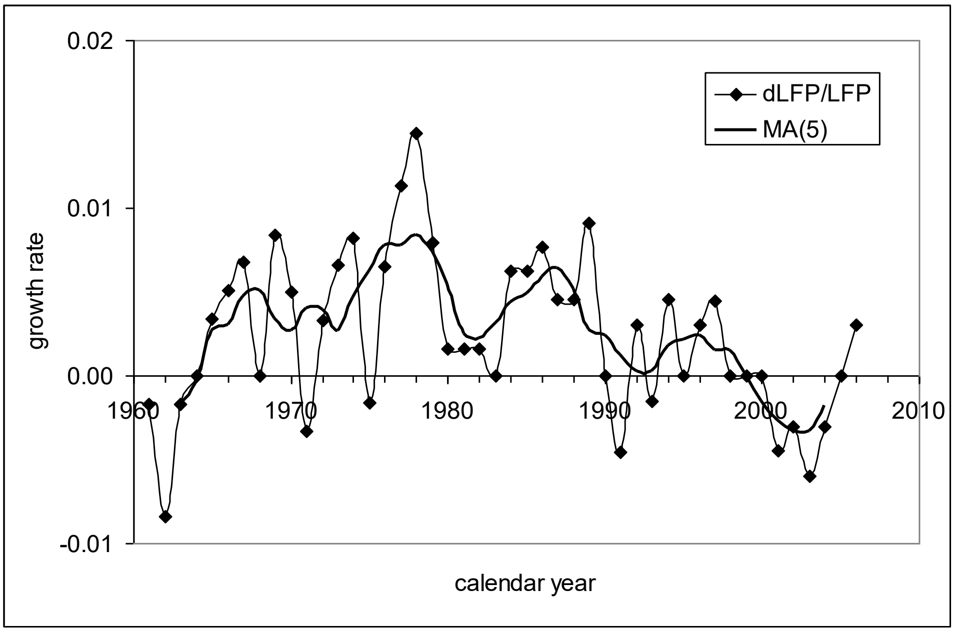


Figure 1. The growth rate of labor force participation (LFP) in the U.S. between 1960 and 2007 as reported by the Bureau of Labor Statistics. The original time series is smoothed by a centered 5-year moving average, MA(5). Notice the increase in the growth rate of the LFP since 2003.

The knowledge of the evolution of labor force participation is crucial for the development of appropriate budgetary, tax, macroeconomic, and financial policy. This is the task for such institutions as the Congressional Budget Office (2004, 2006), the Bureau of Labor Statistic (BLS) (Su, 2007; Toossi, 2005, 2007; Figueroa and Woods, 2007), and the Federal Reserve Banks (FRB San Francisco, 2007; Himmelberg and McConnell, 2005). For example, the CBO (2006)

> … developed a microsimulation approach for analyzing Social Security and other long-term policy issues in order to provide the Congress with comprehensive analyses of the budgetary, distributional, and aggregate economic aspects of various policy choices. The microsimulation approach makes it possible to examine how policy affects individuals' benefits under current law and proposed alternatives, including individual accounts.

These models and corresponding LFP projections are characterized by different forecasting accuracy at various time horizons [CBO, (2004)]. Also, the U.S. Census Bureau (CB) revises the Current Population Survey (CPS) methodology and procedures with possible inconsistencies and incompatibility of data over time [CB, (2002, 2005)]. In 1979, the Census Bureau applied new population controls obtained in the 1980 census and implemented a new questionnaire. In 1994, the CB introduced computer-assisted interviewing techniques with new definitions of the labor force concepts [Polivka, (1996); Polivka and Miller, (1998); Polivka and Rothgeb, (1993)]. The redesigned CPS reports more individuals as being in labor force than did the old survey. This effect is an age-, race-, and gender-dependent, however, and produces artificial steps of different height in the time series for various population groups. Such steps result in even higher spikes in the time derivatives of these time series.

Bearing in mind existing economic models for the LFP evolution, numerous factors apparently influencing LFP, and data availability and quality for the U.S. and some other developed countries we have developed a model, which links aggregate LFP to single economic variable – real GDP per capita. This link is fundamental and expresses the inherent trade-off between economic growth and personal income distribution. Our model is quantitatively formulated and tested on the sample of the U.S. and then validated by modeling LFP in other developed countries.

The remainder of the paper is organized as follows. Section 1 presents some working assumptions on quantitative links between labor force participation rate, personal income distribution, the growth rate of real GDP per capita, and the number of 9-year-olds. In Section 2, we test these assumptions and quantitative relationships against actual data and present some predictions of the future evolution of labor force participation rate in the U.S. Section 3 presents results of similar analysis for some other developed counties, which tests and validates the trade-off between real economic growth and labor force participation rate. Section 4 concludes.

2. The model

Our principal assumption consists in the existence of an inherent trade-off between the mechanisms of personal income distribution (PID) and economic growth. In the U.S., personal income distribution has not been changing much since the start of corresponding measurements in 1947 [Kitov, (2007)]. So, the changes in the rate of economic growth have been accommodated by some changes in relative performance of these mechanisms of income redistribution, not by changes in the distribution itself. Obviously, the rate of participation in labor force, as one of such mechanisms of income redistribution, has been changing over time. Therefore, it is reasonable to start with some features of the PID in the U.S.

The distribution of personal incomes in the U.S. has two branches – quasi-exponential one for incomes from zero to some level, which is called the Pareto threshold. From this threshold, the personal incomes are distributed according to a power law or the Pareto law. Kitov (2005ab) showed that each and every personal income, except the highest ~10% characterized by the Pareto distribution, grows with an approximately constant annual increment to some critical work experience [Kitov, (2005c)]. This quasi-constant increment leads to an exponential growth (of average income for given age) with a negative index: *{1-exp(-at)},* where *a* is a small index, *t* is the working experience. This is a type of satiation process for personal income with age. The process of the average income growth stops when the working experience reaches some critical value, T_{cr}. This critical working experience evolves with time as the square root of real GDP per capita, as shown by Kitov (2005c). Then, an exponential decay of the average income with work experience is observed. For our purposes, the most important empirical fact is that personal income distribution in the U.S. has been practically not changing, when normalized to contemporary working age population and total personal income [Kitov, (2007)]. Effectively, fractions of total personal income have been distributed in the same way among the same fractions of the working age population. As a result, the Gini coefficient for the personal incomes reported by the CB has not been changing over time.

Let's assume that all persons who have a paid job or its equivalent do participate in the production of real GDP. Similarly to the personal income distribution, personal inputs to the real GDP should also be distributed over working population according to some functional dependence on the input. It would be not too inaccurate to assume that these personal inputs to the GDP are exponentially distributed, i.e. the number of people with given input increases exponentially with a decreasing level of the input.

What are qualitative and quantitative effects of real economic growth on LFP considering the fixed (normalized) PID and relevant inputs to the GDP in the U.S.? As shown by Kitov (2006cd), increasing labor supply has no effect on the growth rate of real GDP per capita since the latter depends only on the attained level of real GDP per capita, *G*, and on the changing number of 9-year-olds, *N9*, according to the following relationship:

$$dG/G = 0.5dN_9/N_9 + A_1/G \qquad (1)$$

or, in the reversed form,

$$dN_9/N_9 = 2(dG/G - A_1/G) \qquad (1')$$

where A_1 is an empirical constant. The term A_1/G can be associated with potential economic growth. In other words, real economic growth in the U.S. would be inversely proportional to the attained level of real GDP per capita if the N_9 has not been changing.

Figure 2 compares the number of 9-year-olds measured by the Census Bureau and that predicted by (1′), where the estimates of real GDP per capita are borrowed from the Conference Board database (2008). (For the U.S. and other developed countries in this study, GDP per capita in 1990 U.S. dollars is used, i.e. that converted at Geary Khamis PPPs.) Our model explains 80% (R^2 =0.8) of the variability in real GDP per capita between 1960 and 2006. This is an excellent result considering the accuracy of measurement of both G and N_9. (Notice that the prediction of real economic growth at a nine-year horizon is natural.)

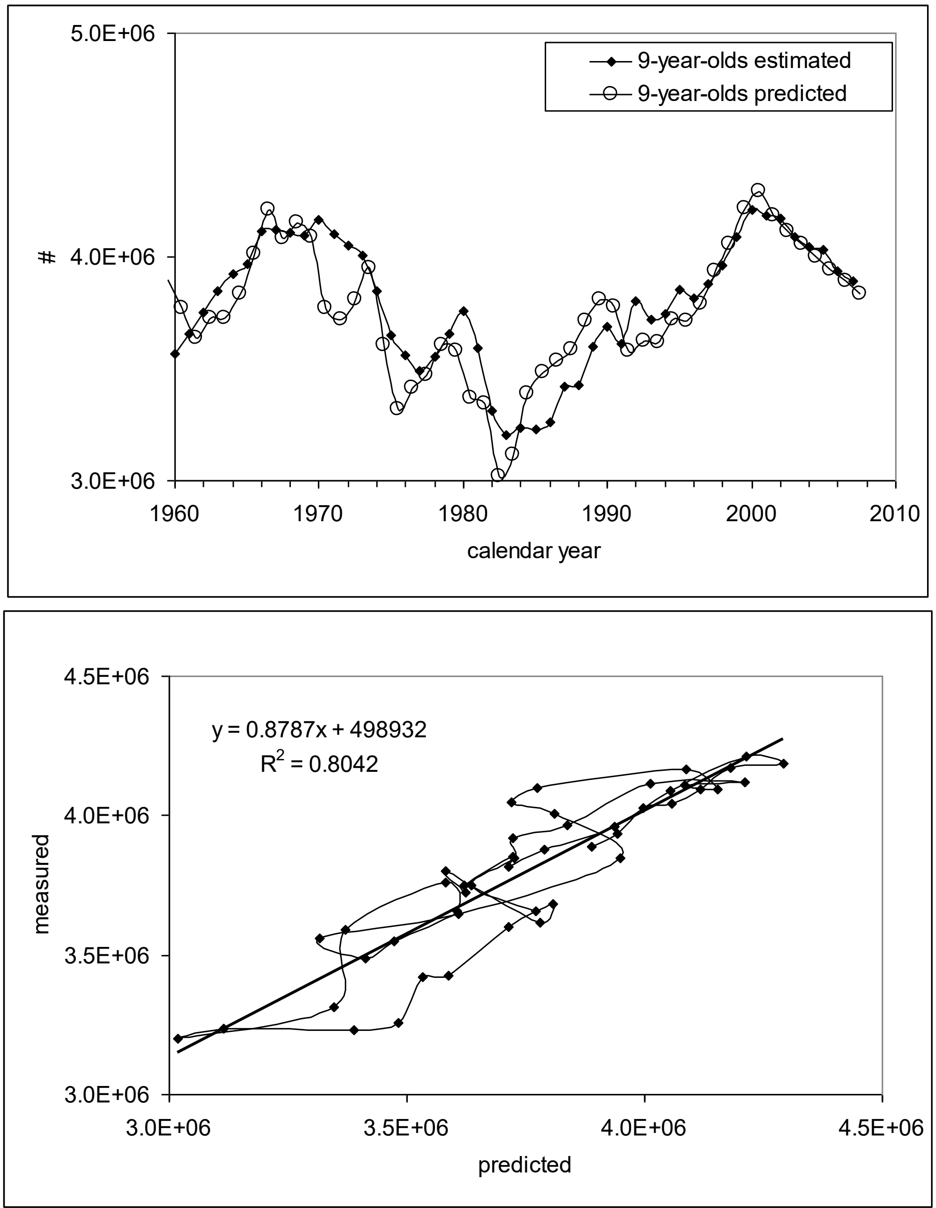


Figure 2. Observed and predicted number of 9-year-olds in the U.S. The predicted number is obtained from the estimates of real GDP per capita according to (1) with A_1=$398 (1990 U.S. dollars). Notice that this relationship demonstrates a non-linear link between the number of 9-year-olds and the growth rate of real GDP per capita. It is presumed in this study that the latter variable is the driving force of the growth in labor force participation rate. Linear regression is also depicted in the lower panel and provides an estimate of the goodness-of-fit: R^2 = 0.8.

On the other hand, the growth in real GDP can influence labor force supply through redistribution of personal incomes. Fluctuations in the number of 9-year-olds produce fluctuations in real GDP per capita relative to that defined by potential economic growth, A_1/G, and thus provide variations in personal income relative to some neutral growth rate. The simplest assumption on the redistribution of some excessive (positive) amount of personal income consists in the increase of the fraction of population in labor force. At first glance, more people would be able to obtain paid jobs with extra money in the economy. Surprisingly, this assumption is wrong for the U.S. The intuition behind the mechanism of LFP reaction on the redistribution of the money excess is opposite – less people are forced to seek income through paid job because of other channels (not included in the CPS questionnaire) of personal income redistribution. A smaller part of the working age population receives more income and somehow transfers it to the fraction of the populations not in labor force to recover original PID. When the growth rate of real GDP per capita is below its potential the total personal income grows at a rate below the neutral one and the lack of personal income earned by given labor force has to be compensated by an increase in the LFP. Figure 2 demonstrates that the N_9 was on a downward trend in the late 1960s and the1970s. Hence, these years are characterized by the growth rate of real GDP per capita below its potential and, thus, by an increasing labor force participation. On the contrary, the years after 1983 have to demonstrate a decreasing rate of the growth in the labor force participation. All these effects can be observed in Figure 1.

The influence of the growth in real GDP on the LFP has to be complicated by the presence of the exponential distribution of personal inputs to real GDP – the number of people with given income roll-off exponentially as a function of income. If the effect of real growth is based on the excess of the total personal income above its potential level, then higher levels of the LFP are more sensitive to real growth. Really, more people can be included in or excluded from the redistribution because of their smaller personal incomes for paid jobs, which are replaced by some other (not measured) mechanisms of personal income earning. It is reasonable to assume that the sensitivity of the LFP to the difference between actual and potential growth rates, $g(t)=dG/G–A_1/G$, increases exponentially with increasing LFP. Also, there might be a time delay between action and reaction and the LFP may lag behind the $g(t)$. Now we are ready for a quantitative analysis with a tentative relationship:

$$\{B_1 dLFP(t)/LFP(t) + C_1\}\exp\{ \alpha_1[LFP(t) – LFP(t_0)]/LFP(t_0) =$$
$$= \int\{dG(t-T))/G(t-T) – A_1/G(t-T)\}dt \qquad (2)$$

where B_1 and C_1 are empirical (country-specific) calibration constants, α_1 is empirical (also country-specific) exponent, t_0 is the start year of modeling, T is the time lag, and $dt=t_2-t_1$, t_1 and t_2 are the start and the end time of the time period for the integration of the $g(t)$ (one year in our model). The exponential term defines the change in the sensitivity due to deviation of the LFP from its initial value $LFP(t_0)$. Effectively, the $LFP(t)$ is a nonlinear function of real economic growth and the trajectory of the $g(t)$ in the past does matter for the attained level of labor force participation.

A simple transformation of (2) using (1) provides another useful form of relationship (2), which relies on the $N_9(t)$ instead of the integral of the $g(t)$:

$$\{B_2 dLFP(t)/LFP(t) + C_2\} \exp\{\alpha_2[LFP(t) – LFP(t_0)]/LFP(t_0)\} = N_9(t-T) \qquad (3)$$

where B_2 and C_2 are empirical constant different from B_1, C_1, and $\alpha_2=\alpha_1$.

Historically, we first tried to model $dLFP/LFP$ as a nonlinear function of G and tested a simple relationship similar to (1):

$$dLFP(t)/LFP(t) = D_1[dG(t-T)/G(t-T) – A_2/G(t-T)] +D_2 \qquad (4)$$

where D_1 and D_2 are empirical (country specific) constants, and A_2 is also an empirical constant, which is different from A_1 in (1). This model served as a workhorse for those countries, which do not provide accurate estimates of the specific age population. Relationship (4) mimics the nonlinear relationship reciprocal to (2). According to (2) one can rewrite (4) in the following (discrete) form:

$$N(t_2) = N(t_1)\{ 2[dG(t_2-T)/G(t_2-T) - A_2/G(t_2-T)] + 1\} \quad (5)$$

$$dLFP(t_2)/LFP(t_2) = N(t_2-T)/B + C \quad (6)$$

where $N(t)$ is the (formally defined) specific age population, as obtained using A_2 instead of A_1, B and C are empirical constants. Relationship (5) defines the evolution of some specific age population, which is different from actual one. The discrete form is useful for calculations. The difference between (3) and (6) consists in the absence of nonlinear terms in (6). It is worth noting that relationships (4) through (6) represent a formal model and do not pretend to fully substitute the meaning of the specific age population, labor force, and real economic growth implied in (2) and (3).

3. Modeling the evolution of labor force participation in the U.S.

As in any empirical analysis, we first discuss availability and quality of relevant data. We use the original source reporting the estimates of labor force participation rate in the U.S. – the BLS (2008). As clear from Figure 1, the original growth rate is characterized by a high volatility induced by relatively low accuracy of corresponding measurements. Therefore, we use MA(5) in our comparisons of observed and predicted time series. One also has to bear in mind that the difference between the LFP measured according to the Census Bureau's definition and some true LFP, which represents the variable for a valid quantitative analysis. Due to numerous revisions to the CB's definition over time this difference is also time dependent and introduces significant noise in the analysis.

Estimates of real GDP per capita can be obtained from many sources. We have chosen the Conference Board database (2008) because it provides PPP estimates for other developed countries what might bring some consistency in the overall cross country comparison. Direct statistical assessment of the uncertainty in the growth rate of the overall real GDP estimates is not available [Fixler and Grimm, (2005)], but numerous revisions to corresponding annual estimates indicate that one percentage point is a reasonable value for the uncertainty.

Considering about the same uncertainty in the estimates of working age population (CBO, 2004) one can conclude that the accuracy of the estimates of the growth arte of real GDP per capita is not better than 1 percentage point as well. Hence, relative differences of 1 to 2 percentage points between measured and predicted values in this study are inside the uncertainty of original estimates and do not need any additional explanation. The above mentioned revisions to population related variables introduce extra uncertainty, however.

Relationship (3) is the start point of our modeling. Visual fit between predicted and measured *dLFP/LFP* has been sought. Figure 3 presents results of the $N_9(t)$ prediction, which can be replaced with G by (2), using the original LFP time series from the BLS.

Corresponding constants are as follows: t_0=1963; T=2 years, □□=-1.85, B_2=-1.5E+8, C_2=4.94E+6. The predicted time series leads the observed one by two years, i.e. an accurate forecast at a two-year horizon is a natural feature of the model. Coefficient B_2 is negative what results in a declining rate of the LFP growth during the years of real growth above the potential one, for example, between 1983 and 2000. Exponential term in (3) provides a factor of 0.77 in 2000 (the largest LFP of 67.1%) relative to 1963, when the LFP was only 58.7%. This means that 1% change in *N9* at the LFP level of 67.1% produces a larger change in the *dLFP/LFP* by factor of 1/0.77=1.3 than 1% change at the level of 58.7%. Effectively, the sensitivity of the *LFP* to the *N9* is 1.3 times larger. Also displayed is the case without exponential weighting, □□=0. This case demonstrates that the specific age population (N_9) is overestimated by the model.

Considering the uncertainty in the underlying time series – N_9 and *LFP,* the observed and predicted time series are in a good overall agreement: timing of main turns in both series is excellent and amplitudes of the largest changes are also practically coincide.

The measured number of 9-year-olds is characterized by a high uncertainty related to two sources – methodology and procedures of measurements in decennial censuses and interpolation of the census estimated between the years of censuses using birth-death-immigration components of the population projection. West and Robinson (1999) estimated the uncertainty in the number of 9-year-olds for the 1990 census between 3% and 5% in the age group between 5 and 9 years. Especially high revisions to the population estimates are made near census years – as seen from Figure 3. The largest deviations between the predicted and observed population are around the censuses.

So, Figures 2 and 3 demonstrate that real economic growth (or N_9, which is an equivalent to G) may explain the evolution of labor force participation rate in the U.S. after 1963. There is another way to invert equations (2) and (3), as represented by relationships (5) and (6). These relationships allow expressing *LFP* as a nonlinear function of *G* what facilitates the prediction of labor supply. At the same time, relationships (5) and (6) are just a useful approximation in the case when necessary population estimates are absent or poor.

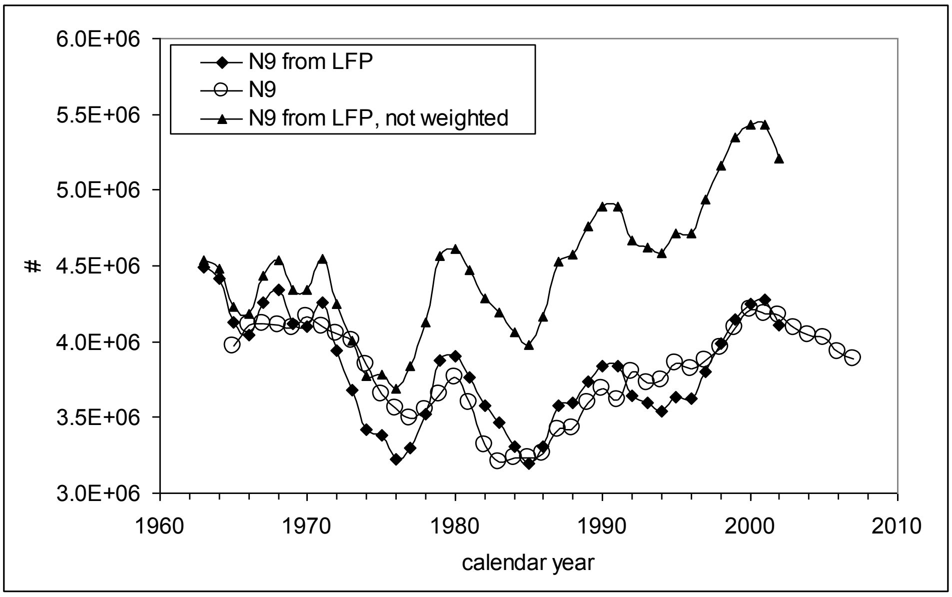
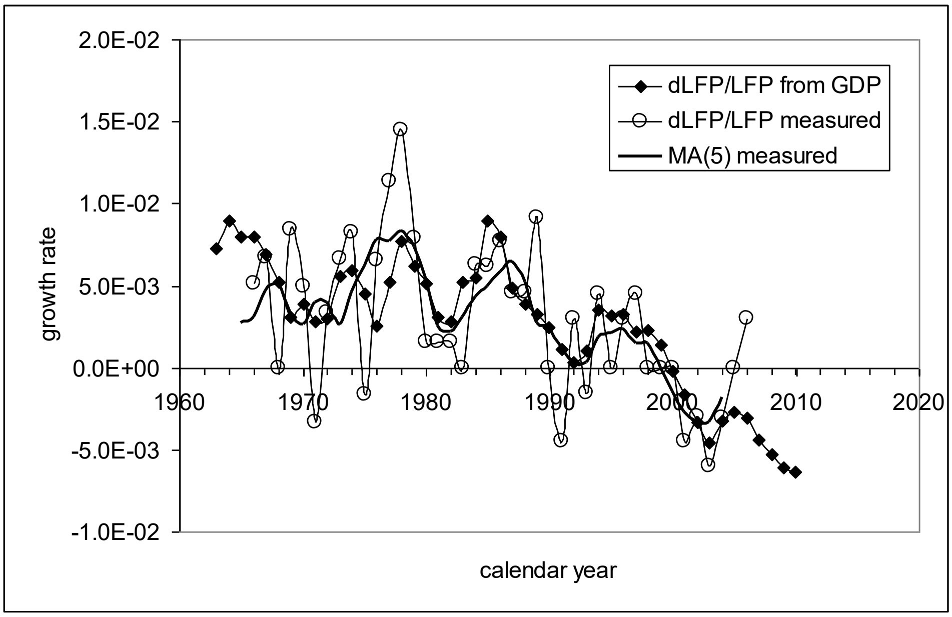


Figure 3. The number of 9-year-olds: the observed one and that obtained from the LFP with and without exponential weighting in (3). Constants t_0=1963; B_2=-1.5E+8, C_2=4.94E+6, α_2=-1.85. Also shown is the case with α_2=0: N_9 is overestimated.



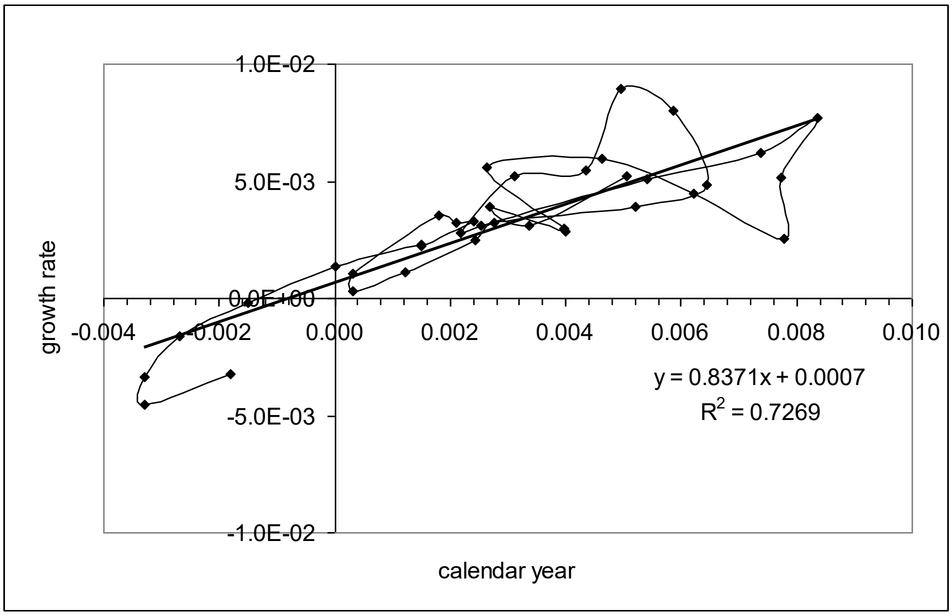


Figure 4. Observed and predicted growth rate of LFP in the U.S. The latter is obtained from real GDP per capita using (5) and (6) with *N(1959)*=4.5E+6, A_2=\$350 (1990-dollars), *B*=-1.23E+8, *C*=0.04225.
The lower panel presents a regression of the observed time series on the predicted with R^2=0.73.

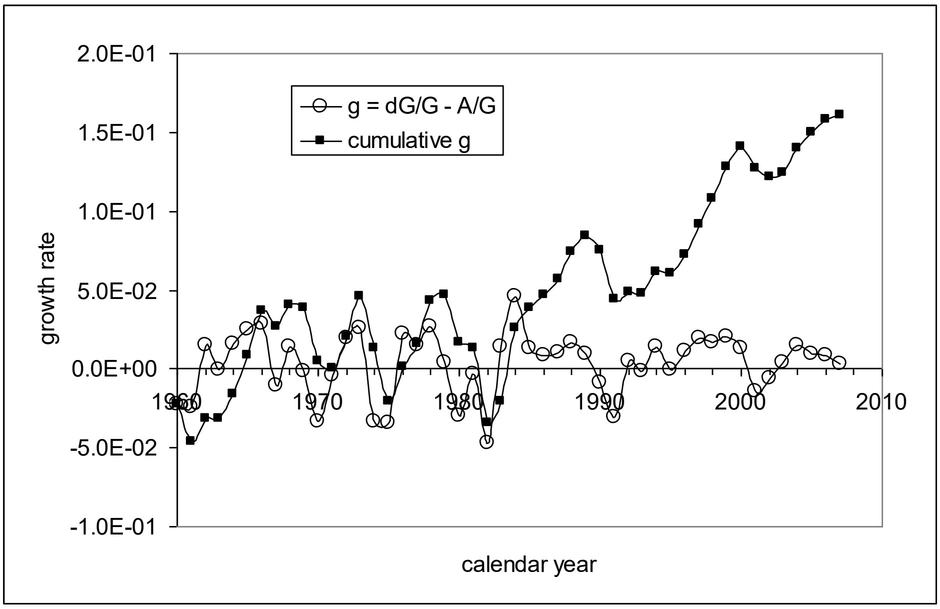


Figure 5. Evolution of the transient part, *g(t)*, of the growth rate of real GDP per capita, *G*:
$g(t) = dG(t)/G(t) – A_2/G(t)$, where A_2= \$350 (1990-dollars) is empirical (and country-specific) constant.
Also shown is the cumulative value of the transient part. Notice the growth of the cumulative value since 1983.

Figure 4 depicts the observed and predicted relative change rate of the *LFP*. The latter is obtained from (5) and (6) with the following constants and coefficients: *N(1959)*=4.5E+6, A_2=\$350, *B*=-1.23E+8, *C*=0.04225. Notice that coefficient A_2 is smaller than A_1=\$398 in (1). Due to high volatility of the original *dLFP/LFP* time series we compare the predicted time series to MA(5) of the observed series – corresponding linear regression is shown in the lower panel of Figure 4. The goodness-of-fit is high: R^2=0.73. More important is that timing of main turns in the observed time series is well predicted at a two-year horizon – the predicted series is still two years ahead of the observed one: T=2 years. Figure 5 displays the evolution of the $g(t)$ and its cumulative value. Since $A_2 < A_1$, the cumulative value of $g(t)$ has a strong positive trend after 1983. Reciprocally, the $dLFP/LFP$ is characterized by a negative trend.

Having the prediction of the $dLFP/LFP$ from the G, one can predict the LFP itself. Figure 6 compares the original labor force participation rate and that obtained from the predicted time series presented in Figure 4. Both curves are synchronized due to a two-year back-shift of the predicted curve. There are three major deviations between the curves – all are associated with large revisions to the original LFP. The root-mean-square forecasting error (RMSFE) of the LFP at a two year horizon for the period between 1968 and 2006 is 0.28%. This accuracy is slightly larger than that inherent to the original time series (from 0.1% to 0.2%) and, noticeably, obtained using raw data. The elimination of the most obvious (artificial) deviations induced by the revisions to the CPS results in RMSFE=0.2%.

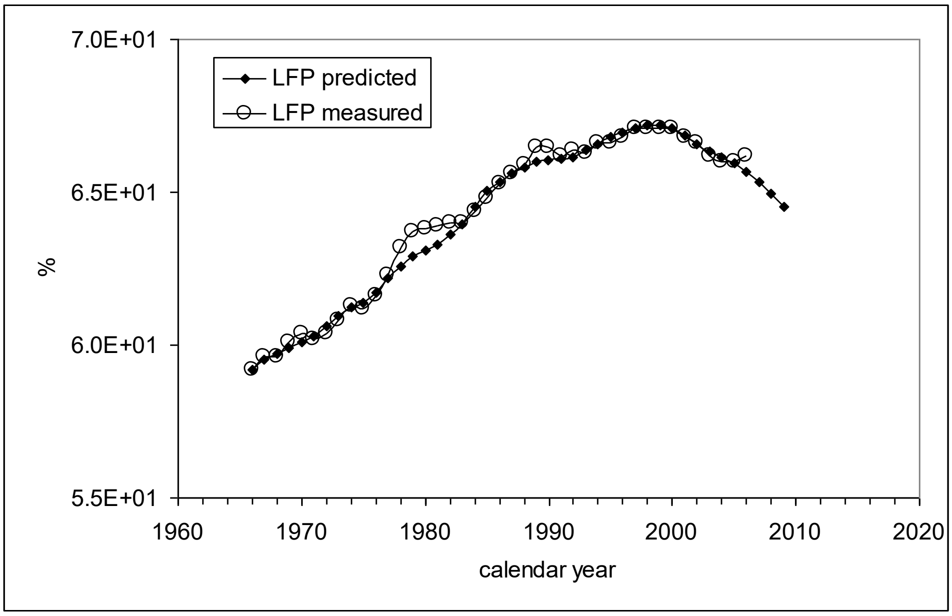
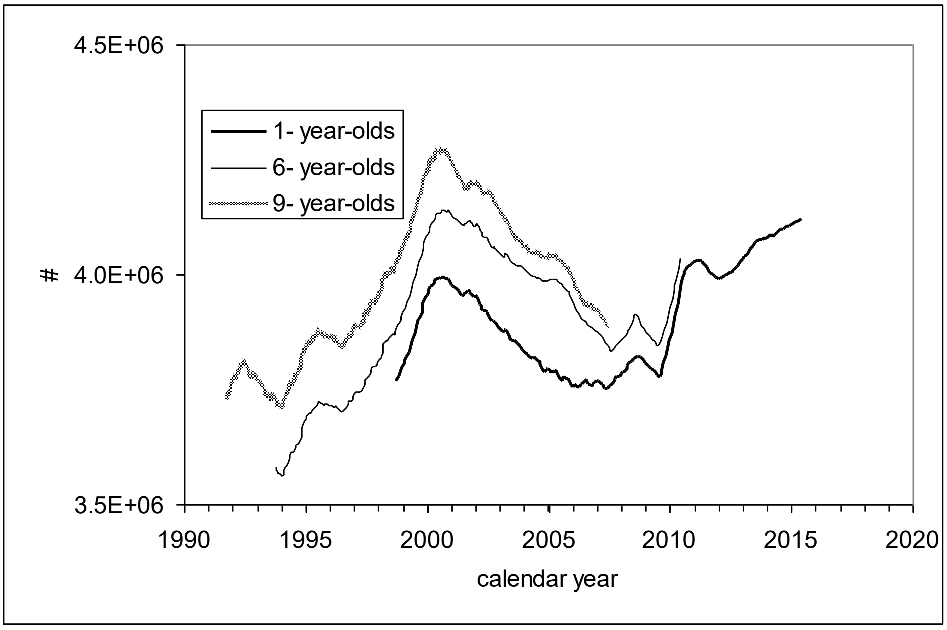


Figure 6. Observed and predicted LFP in the U.S. The latter is obtained from the growth rate time series presented in Figure 5. Notice the largest deviation between the curves is associated with the years of major revisions to the LFP – 1980 and 1990.

Relationship (1) provides a unique opportunity to foresee the evolution of the LFP 11 years ahead. Really, the number of 9-year-olds can be predicted from the younger birth cohorts and additional two years are given by the lag of the LFP behind real economic growth.

Figure 7 shows qualitatively what will be the direction of the LFP evolution. It displays the projections of N_9 and dN_9/N_9 using 6- and 1-year-olds, as available by the end of 2007. The years 2007 through 2010 are the turning ones for real economic growth – the negative trend observed between 2000 and 2009 should be switched to a positive one. But even poor years for the U.S. economy between 2001 and 2010 are characterized by a negative trend in the $dLFP/LFP$, as Figure 4 demonstrates. This effect is related to the difference between A_1 and A_2 ($A_1 > A_2$) – even real economic growth at a rate slightly below its potential value A_1/G is still above the A_2/G, which is the watershed for the growth rate of the LFP. Really nasty years, i.e. those with very low or even negative dG/G, are necessary to change the negative ($dLFP/LFP$) trend to a positive one.

a.



b.

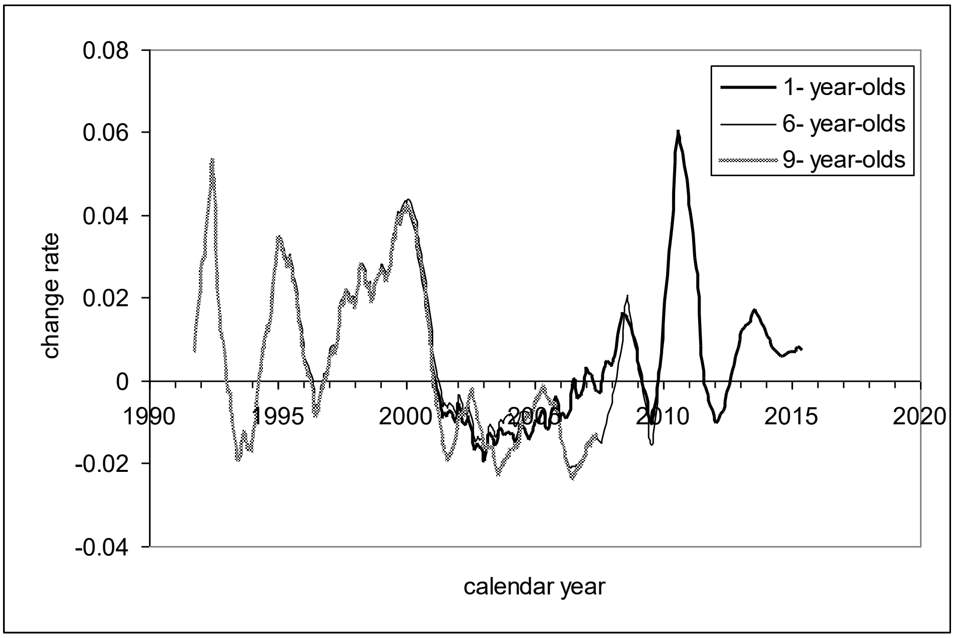


Figure 7. Prediction of the number of 9-year-olds by extrapolation of population estimates for younger ages (1- and 6-year-olds). A) Total population estimates. The time series for younger cohorts are shifted ahead by 8 and 3 years, respectively. B) The change rate of the population estimates given in a), which is proportional to the growth rate of real GDP per capita. Notice the difference in the change rate provided by 1-year-olds and 6-year-olds for the period between 2003 and 2010. This discrepancy is related to the age-dependent difference in population revisions.

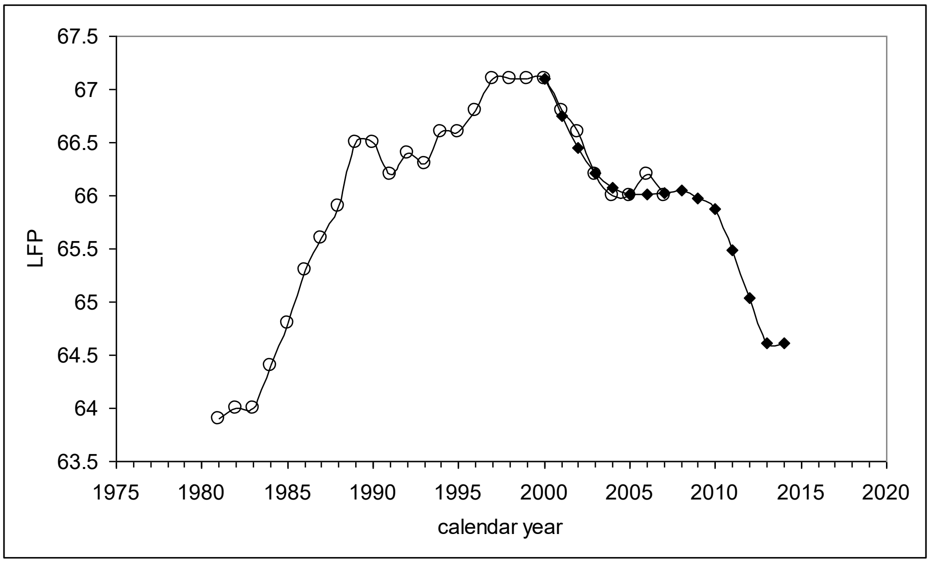


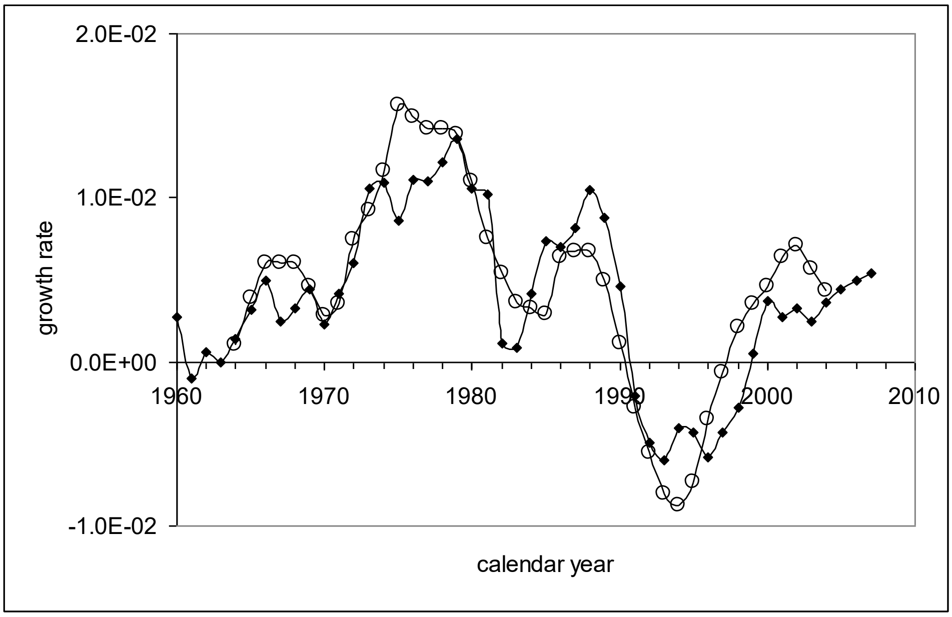
Figure 8. Prediction of the LFP evolution in the USA between 2000 and 2014 from the number of 3-year-olds. Flat segment between 2004 and 2009 will end up in a rapid drop by 1.3% after 2010. This is the effect of an elevated (above potential) real economic growth.

The acceleration of real economic growth, which will be observed after 2010, will also bring additional acceleration to the process of the declining labor force participation rate, as Figure 8 demonstrates. The effect of the declining LFP will be mapped into a slowdown in the growth in labor force itself. As a consequence, inflation will drop close or below zero [Kitov, (2006ef); Kitov, Kitov, Dolinskaya, (2007a)]. All in all, the Golden Era will come back – real economic growth above its potential, low inflation, and low participation rate – the middle class has enough money to run families. What could be better?

4. Modeling the evolution of LFP in developed countries

There are two general methods to validate empirical relationships – to extend data set for one system or to describe with the same model the evolution of similar systems over the same time interval. In Section 2, we demonstrated the accuracy of our empirical model linking LFP and real economic growth in the U.S. for the period between 1963 and 2006. In this Section, we extend our approach and model LFP to some other developed countries. One can consider this study as a validation of the model for the U.S. Therefore, we just present quantitative results not discussing them in details.

We begin with the closest neighbor of the U.S. – Canada. The upper panel in Figure 9 depicts the observed and predicted rate of LFP change. The predicted time series is obtained using the following coefficients in (5) and (6): $N(1959)$=270000, A_2=\$342, B=4.0E+6, C=-0.0607. We did not use linear regression to obtain these coefficients. In order to obtain the best fit, only visual similarity between cumulative values was sought. In many aspects, the approach based on cumulative representation provides a useful constraint on the coefficients in relationships (5) and (6) due to high sensitivity of cumulative values to small disturbances and deviations (Kitov, Kitov, Dolinskaya, 2007ab). This constraint is similar to mass and energy conservation law in physics. Regression usually underestimates coefficients in linear relationships. The lower panel in Figure 9 presents the measured LFP and that obtained from the predicted *dLFP/LFP* in the upper panel. One can observe an excellent agreement in timing and amplitudes of the changes in both LFPs – mainly these curves coincide. This agreement supports our assumption that real economic growth completely defines the evolution of LFP in developed countries. The difference between the cumulative curves is likely defined by measurement noise. Therefore, no other factors are needed for the description of the observed LFP.



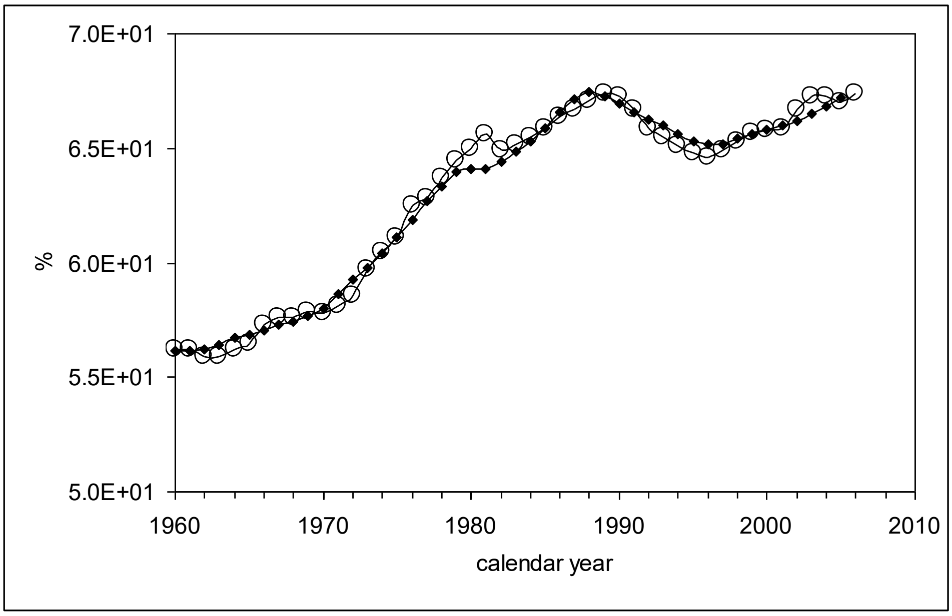


Figure 9. Upper panel: observed [MA(5)] and predicted growth rate of LFP in Canada: *N(1959)*=270000, A_2=\$342 (1990 U.S. dollars), *B*=4.0E+6, *C*=-0.0607, *T*=0 years. For the period between 1963 and 2006, a linear regression gives the goodness-of-fit, R^2=0.78.

The lower panel depicts the observed and predicted LFP.

Canada provides one of the best examples of successful modeling for the entire period between 1960 and 2006 and s also characterized by a wide *dLFP/LFP* dynamic range: from -1.0E-2 [y^{-1}] to +2.0E-2 [y^{-1}], with the LFP changing from 55.9% in 1963 to 67.4% in 2006. Dynamic range is a crucial characteristic for any empirical study because it defines resolution. The wider is the dynamic range the higher is the resolution, for the same signal to noise ratio. The noise is defined by the uncertainty in corresponding measurements. As mentioned above, cumulative (or integral) values significantly improve the resolution due to larger incoming signals.

The modeling of the evolution of labor force participation in Italy has also been successful. The upper panel in Figure 10 displays the measured and predicted *dLFP/LFP*. The later is obtained using the following coefficients: *N(1959)*=570000, A_2=\$270; *B*=1.3E+7, *C*=-0.0667, *T*=0. (Notice a relatively low value of A_2.) Linear regression of the time series in the upper panel (between 1963 and 2006) is characterized by R^2=0.71.

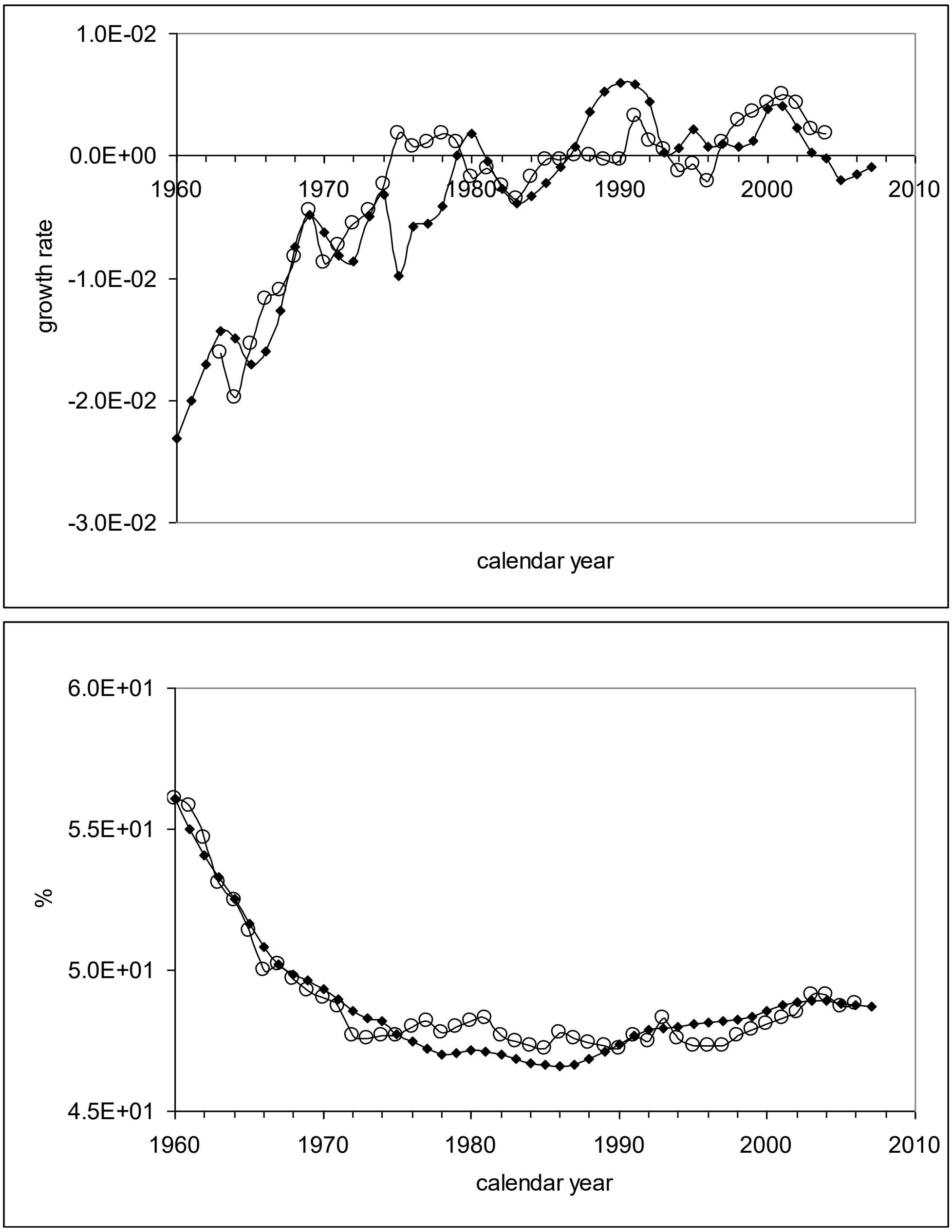


Figure 10. Observed [MA(5)] and predicted growth rate of LFP in Italy: *N(1959)*=570000, A_2=\$270 (1990 U.S. dollars), *B*=1.3E+7, *C*=-0.0667, *T*=0. For the period between 1963 and 2006, a linear regression gives the goodness-of-fit, R^2=0.71.

The lower panel depicts the original LFP, changing in the range from 56.1% in 1960 to 47.2% in 1984, and the predicted LFP.

The lower panel in Figure 10 shows that the dynamic range of LFP for Italy is also wide – from 56.1% in 1960 to 47.2% in 1984. The predicted cumulative curve is very close to the measured one over the entire period between 1960 and 2006. This is a clear indication that real economic growth completely defines the evolution of labor force participation in Italy. Small deviations are potentially induced by measurement noise and revisions to labor force definitions.

France is an opposite example of a country with a narrow range of LFP change and, correspondingly, low growth rate, *dLFP/LFP*. The upper panel in Figure 11 indicates that absolute value of the change rate is generally below 0.003 [y $^{-1}$] between 1963 and 2000. The LFP, as the lower panel in Figure 11 shows, dropped quickly between 1960 and 1963 – from 59.5% to 57%. This period is described relatively well with our model. The period between 1997 and 2005, is characterized by an increase in the LFP from 55.6% to 57%. We failed to simulate this period with our model, but the deviation between the observed and predicted rates are relatively small: ~0.005 [y $^{-1}$]. The cumulative curve related to the predicted *dLFP/LFP* also deviates from the observed LFP. Due to the narrow dynamic range, it is difficult to associate this deviation with one principal source. This deviation does not deny the model, however.

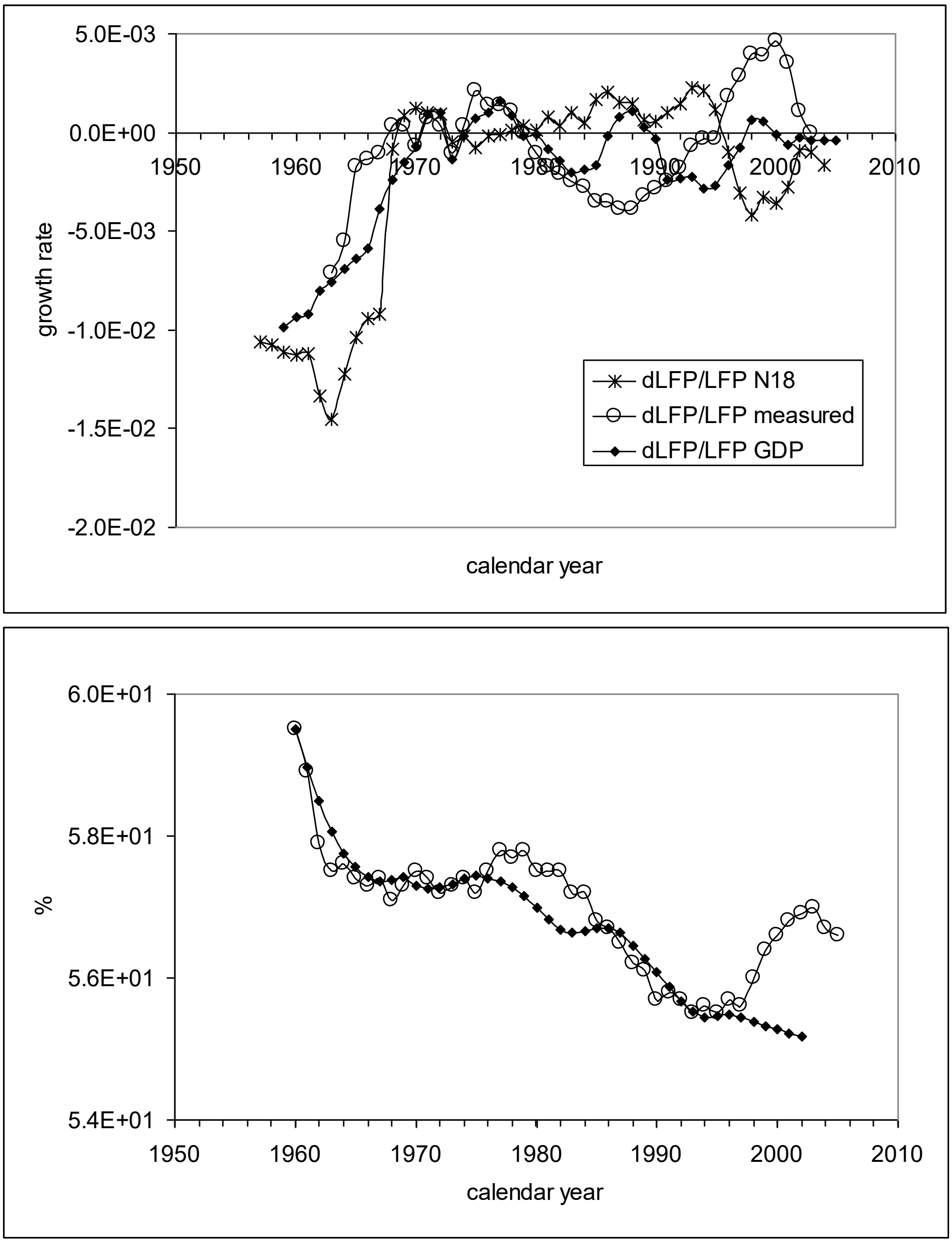


Figure 11. Observed and predicted growth rate of LFP in France: *N(1959)*=570000*, A*=$290 (1990 U.S. dollars)*, B*=2.0E+7*, C*=-0.0425, *T*=0. For the case with *N18*: *B*=2.4E+7, *C*=-0.035.

The lower panel depicts the original LFP, changing in the range from 59.5% in 1960 to 55.5 % in 1995, and the predicted LFP.

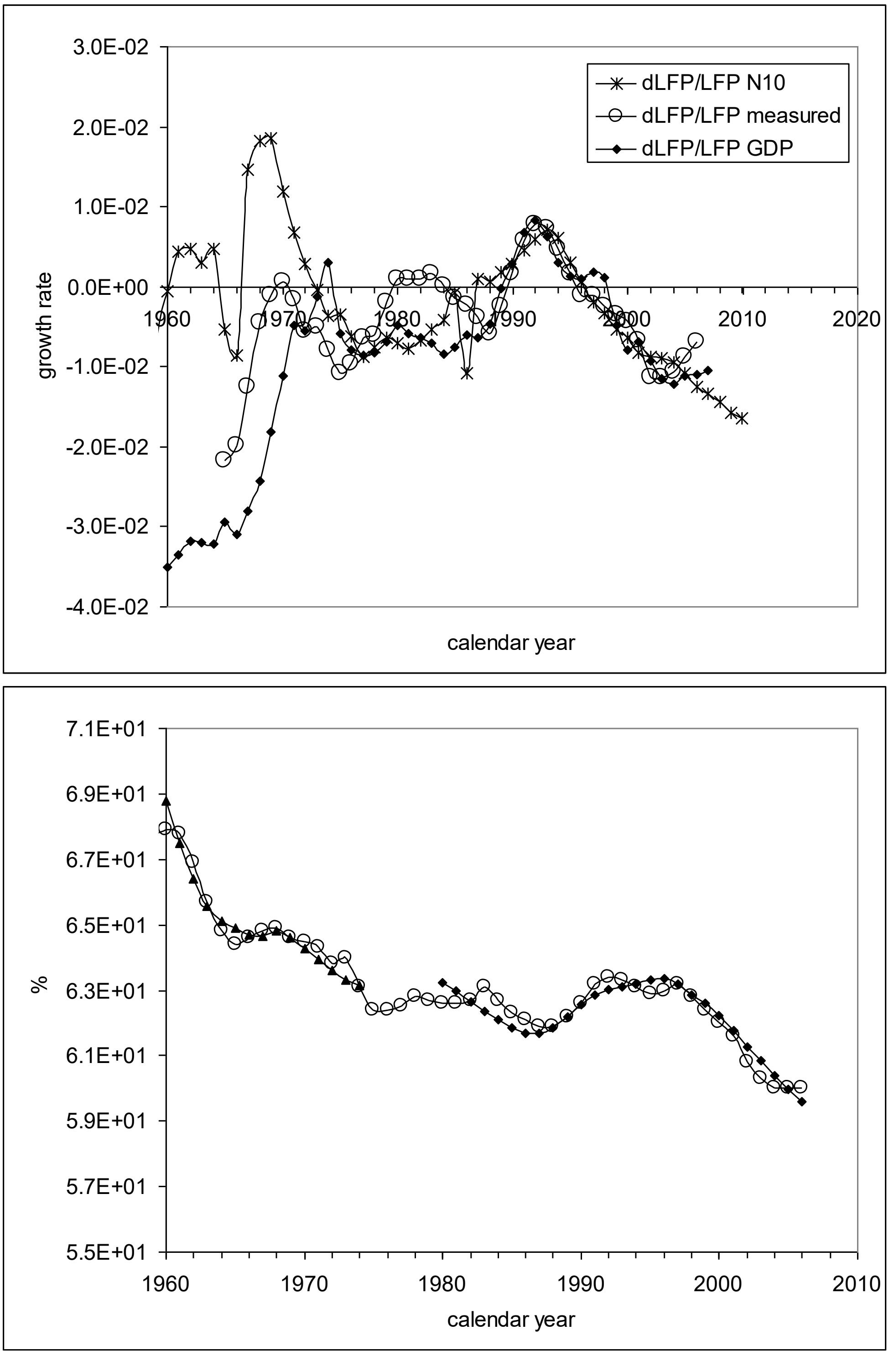
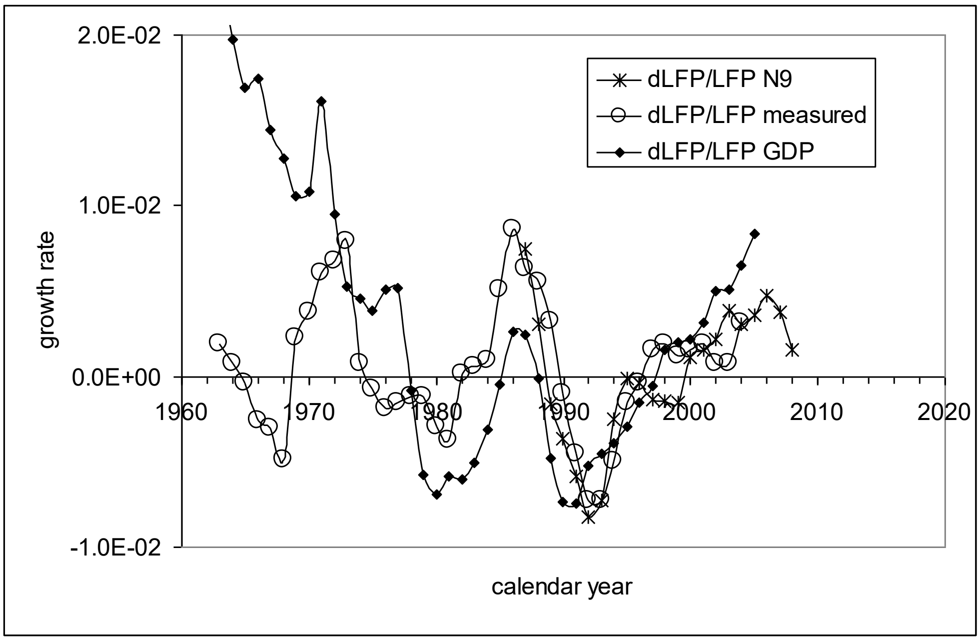


Figure 12. Observed and predicted growth rate of LFP in Japan: *N(1959)*=1570000*, A*=\$390 (1990 U.S. dollars)*, B*=6.5E+7*, C*=-0.048, *T*=0. For the case with *N18*: *B*=3.5E+7, *C*=-0.052.

The lower panel depicts the original LFP, changing in the range from 68% in 1960 to 61% % in 2005, and the predicted LFP. The latter consists of two separated segments. We failed to predict LFP between 1972 and 1978.



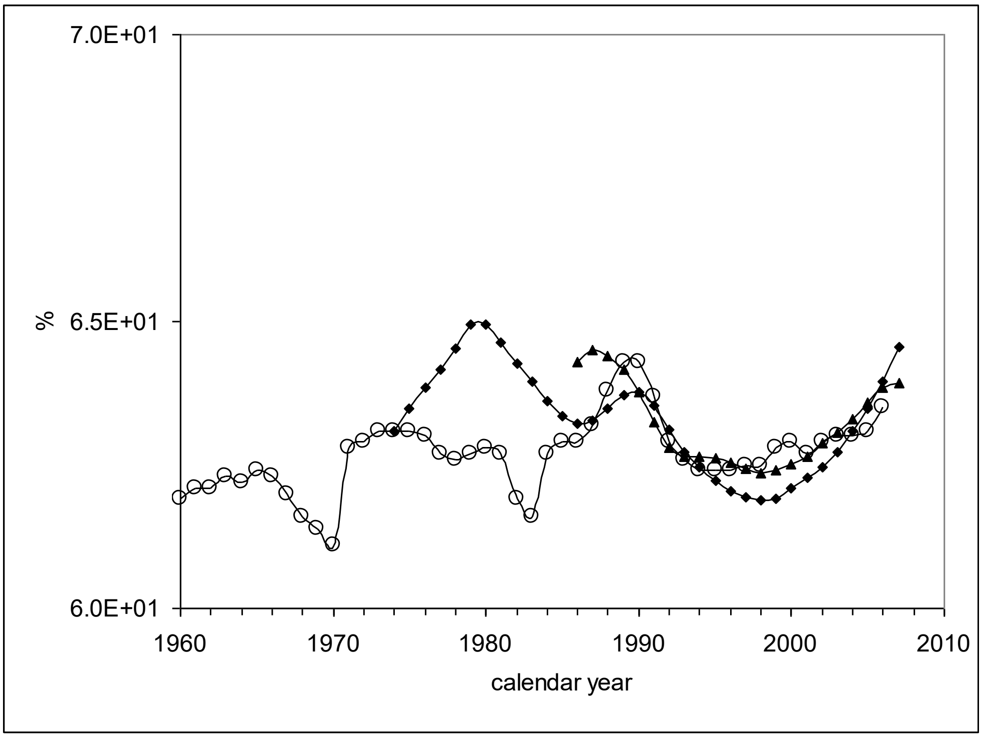


Figure 13. Observed and predicted growth rate of LFP in the UK: *N(1959)*=570000, A_2=\$340(1990 U.S. dollars), *B*=6.0E+7, *C*=-0.062, *T*=0.

For the case with *N9*: *B*=1.15E+7, *C*=-0.064.

The lower panel depicts the original LFP, changing in the range from 61% in 1970 to 64.5 % in 1990, and the predicted LFP.

Japan and the UK, presented in Figure 12 and 13, respectively, are very similar in an accurate prediction of *dLFP/LFP* from both *G* and the specific age population: 18 years of age for Japan and 9-years of age for the UK [Kitov, (2006bd)]. Coefficients for corresponding models are given in figure captions. These accurate predictions, however, span a narrower period after ~1970. The predicted curves describe amplitudes and timing of major turns in the observed curves. This is a principal feature because conventional LFP models are based on the extrapolation of existing (mainly linear) trends and do not foresee any turns. Our model allows prediction of the LFP evolution in Japan at 18-year horizon! The discrepancy before 1970 are not well explained and might be linked to revisions to labor force and real economic growth definitions, and measurement errors.

Labor force participation in Sweden has been also well described after 1975, as Figure 14 shows. It is similar to Japan and the UK but there are no independent estimates provided by specific age population. The dynamic range of the *dLFP/LFP* changes is also not specifically wide – from -0.008 [y $^{-1}$] to +0.002 [y $^{-1}$]. Cumulative curves are very close in the lower panel of Figure 14.

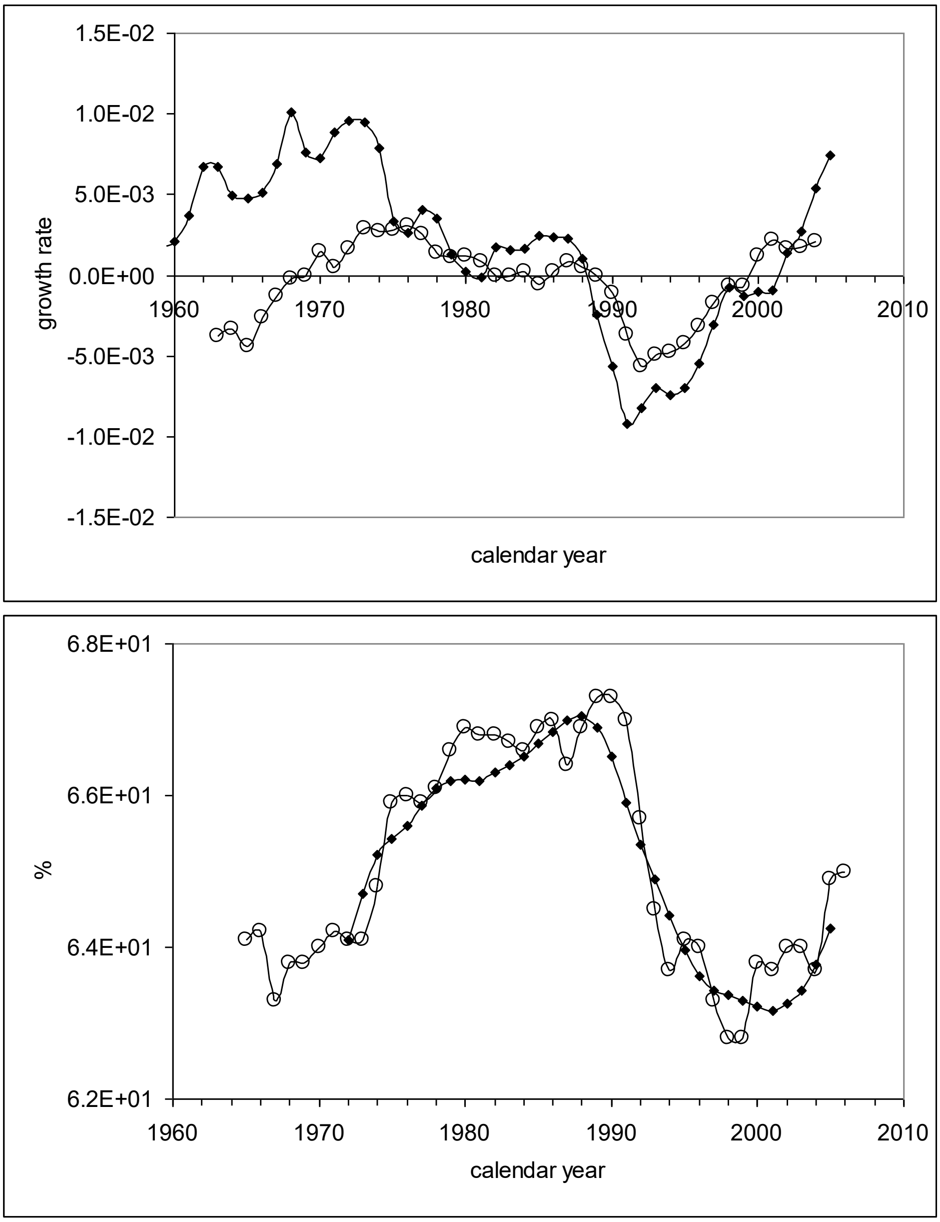


Figure 14. Observed and predicted growth rate of LFP in Sweden: *N(1959)*=100000, A_2=\$310 (1990 U.S. dollars), *B*=2.2E+6, *C*=-0.0465, *T*=0.

The lower panel depicts the original LFP, changing in the range from 67% in 1990 to 62.5 % in 1998, and the predicted LFP.

Six countries were analyzed and all demonstrated results positive for the validation of our model with real economic growth driving labor force participation. Apparently, more investigations are necessary including revisions to data compatibility and quality. A straightforward extension would be the inclusion of other developed countries. There is also a mystery associated with the fact that the effect of real growth on LFP in all studied countries is opposite to that in the U.S., where coefficient *B* is negative.

5. Conclusion

There exist a trade-off between real economic growth and the evolution of labor force in developed countries. Moreover, the results obtained in this study for the period between 1960 and 2007 quantitatively support the assumption that real GDP per capita is likely to completely define LFP.

One fundamental difference between the U.S. and other countries consists in the opposite signs of coefficient *B*: negative for the U.S. and positive for all other studied countries. Effectively, the U.S. is the only country where an elevated real economic growth results in decreasing labor force participation. Such a striking contrast in labor force behavior needs a special analysis.

The evolution of the aggregate LFP puts a strong constrain on the evolution in various age-gender-race groups. It is likely that the mechanism of the influence of real economic growth on LFP in these groups is essentially the same. Therefore, the poorest (in average) age groups with the lowermost LFP are less sensitive to real economic growth - youngest and eldest people are among them.

In any case, the obtained empirical relationships are useful in predictions of LFP in developed countries. The prediction of a drop in the LFP in the U.S. in 2010 can significantly contribute to the model validation.

6. References

[1] Aaronson, S., Fallick, B., Figura, A., Pingle, J.,Wascher, W., (2006), *The Recent Decline in the Labor Force Participation Rate and Its Implications for Potential Labor Supply*, Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution, vol. 37(2006-1), pp. 69 – 154.

[2] Aaronson, D., Park, K.-H., Sullivan, D., (2006), *The Decline in Teen Labor Force Participation. Economic Perspectives*, Federal Reserve Bank of Chicago, 30(1), pp. 2 – 18.

[3] Aaronson, D., Park, K.-H., Sullivan, D., (2007), *Explaining the decline in teen labor force participation*, Chicago Fed Letter, No. 234.

[4] Bureau of Labor Statistics, (2008), *Employment and Unemployment. Table - Labor Force Statistics from the Current Population Survey*, retrieved from http://www.bls.gov/cps/cpsatabs.htm, March 30, 2008.

[5] Census Bureau, (2008), *Population estimates*, retrieved from http://www.census.gov/popest/estimates.php, March 30, 2008.

[6] Conference Board, (2008), *Total Economy Database*, retrieved from http://www. conference-board.org/economics/database.cfm, March 30, 2008.

[7] Congressional Budget Office, (2004), *Projecting Labor Force Participation and Earnings in CBO's Long-Term Microsimulation Model*, Background Paper, October.

[8] Congressional Budget Office, (2006), *CBO's Projections of the Labor Force*, September.

[9] Fallick, B., Pingle, J., (2007), T*he Effect of Population Aging on Aggregate Labor Supply in the U.S., Boston FRB*, May, http://www.bos.frb.org/labor2007/papers/pingleII.pdf.

[10] Figueroa, E., Woods, R., (2007). *Industry output and employment projections to 2016,* Monthly Labor Review, November, pp. 53 – 85.

[11] Fixler, D., Grimm, B., (2005), *Reliability of the NIPA Estimates of U.S. Economic Activity*, BEA, http://www.bea.gov/bea/ARTICLES/2005/02February/0205_NIPAs.pdf.

[12] FRB San Francisco, (2007), *Labor Force Participation and the Prospects for U.S. Growth*, Economic Letter, No. 2007-33, November 2.

[13] Fullerton, H.R., Jr., (2003), *Evaluating the BLS labor force projections to 2000,* Monthly Labor Review, October, pp. 3–13.

[14] Hausman, J. A., (1986), *Taxes and Labor Supply. In Handbook of Public Economics*, v. 1, edited by A. Auerbach and M. Feldstein, North-Holland.

[15] Haveman, R., de John, P., Wolfe, B., (1991), *Disability Transfers and the Work Decision of Older Men*. The Quarterly Journal of Economics, 106(3), pp. 939 – 49.

[16] Himmelberg, C., McConnell, M.M., (2005), *How Much Labor Market Slack: The Importance of Adjusting for Trend Participation Rates*. Unpublished paper. Federal Reserve Bank of New York (January).

[17] Hotchkiss, J.L., (2004), *Employment Growth and Labor Force Participation: How Many Jobs Are Enough?* Federal Reserve Bank of Atlanta, Working Paper 2004-25, October.

[18] Hotchkiss, J.L., (2005), *What's Up with the Decline in Female Labor Force Participation?* Federal Reserve Bank of Atlanta, Working Paper 2005-18, August.

[19] Juhn, C., Potter, S., (2006), *Changes in Labor Force Participation in the United States*, Journal of Economic Perspectives, American Economic Association, vol. 20(3), pages 27 – 46.

[20] Kitov, I., (2005a), *A model for microeconomic and macroeconomic development*, Working Papers 05, ECINEQ, Society for the Study of Economic Inequality.

http://ideas.repec.org/p/inq/inqwps/ecineq2005-05.html

[21] Kitov, I., (2005b), *Evolution of the personal income distribution in the U.S.: High incomes*, Working Papers 02, ECINEQ, Society for the Study of Economic Inequality. http://ideas.repec.org/p/inq/inqwps/ecineq2005-02.html

[22] Kitov, I., (2005c), *Modelling the average income dependence on work experience in the U.S. from 1967 to 2002*, Working Papers 11, ECINEQ, Society for the Study of Economic Inequality. http://ideas.repec.org/p/inq/inqwps/ecineq2005-11.html

[23] Kitov, I., (2006a), *Modelling the age-dependent personal income distribution in the U.S.*, Working Papers 17, ECINEQ, Society for the Study of Economic Inequality. http://ideas.repec.org/p/inq/inqwps/ecineq2006-17.html

[24] Kitov, I., (2006b), *The Japanese economy*, MPRA Paper 2737, University Library of Munich, Germany. http://ideas.repec.org/p/pra/mprapa/2737.html

[25] Kitov, I., (2006c), *Real GDP per capita in developed countries*, MPRA Paper 2738, University Library of Munich, Germany. http://ideas.repec.org/p/pra/mprapa/2738.html

[26] Kitov, I., (2006d), *GDP growth rate and population*, Working Papers 42, ECINEQ, Society for the Study of Economic Inequality. http://ideas.repec.org/p/inq/inqwps/ecineq2006-42.html

[27] Kitov, I., (2006e), *Inflation, unemployment, labor force change in the USA*, Working Papers 28, ECINEQ, Society for the Study of Economic Inequality, http://ideas.repec.org/p/inq/inqwps/ecineq2006-28.html

[28] Kitov, I., (2006f), *Exact prediction of inflation in the USA*, MPRA Paper 2735, University Library of Munich, Germany, http://ideas.repec.org/p/pra/mprapa/2735.html

[29] Kitov, I., (2007), *Modeling the evolution of Gini coefficient for personal incomes in the USA between 1947 and 2005*, Working Papers 67, ECINEQ, Society for the Study of Economic Inequality, http://ideas.repec.org/p/inq/inqwps/ecineq2007-67.html

[30] Kitov, I., Kitov, O., Dolinskaya, S., (2007a), *Inflation as a function of labor force change rate: cointegration test for the USA*, MPRA Paper 2734, University Library of Munich, Germany, http://ideas.repec.org/p/pra/mprapa/2734.html

[31] Kitov, I., Kitov, O., Dolinskaya, S., (2007b), *Modelling real GDP per capita in the U.S.: cointegration test*, MPRA Paper 2739, University Library of Munich, Germany. http://ideas.repec.org/p/pra/mprapa/2739.html

[32] Murphy, K.M., Topel, R.H., (1987), *The Evolution of Unemployment in the United States: 1968-1985*. In NBER Macroeconomics Annual 1987. Edited by S. Fischer. The MIT Press.

[33] Polivka, A.E, (1996), *Data Watch: The Redesigned Current Population Survey*. Journal of Economic Perspectives, 10(3), pp. 169 – 80.

[34] Polivka, A.E. , Miller, S.M., (1998), *The CPS After the Redesign: Refocusing the Economic Lens*. In Labor Statistics and Measurement Issues, edited by J. Haltiwanger, M. Manser, and R. Topel. University of Chicago Press.

[35] Polivka, A.E., Rothgreb, J.M., (1993), *Overhauling the Current Population Survey: Redesigning the Questionnaire*. Monthly Labor Review, 116(September), pp. 10 – 28.

[36] Su, B., (2007), *The U.S. economy to 2016: slower growth as boomers begin to retire*, Monthly Labor Review, November, pp. 13 – 32.

[37] Toossi, M., (2005), *Labor Force Projections to 2014: Retiring Boomers*. Monthly Labor Review 128(November), pp. 25 – 44.

[38] Toossi, M., (2007). *Labor force projections to 2016: more workers in their golden years*, Monthly Labor Review, November, pp. 33–52.

[39] U.S. Census Bureau, (2002), *Technical Paper 63RV: Current Population Survey - Design and Methodology*, issued March 2002. Retrieved February 26, 2007 from http://www.census.gov/prod/2002pubs/tp63rv.pdf.

[40] U.S. Census Bureau, (2005), *Changes in Methodology for the March Current Population Survey*, Retrieved February 26, 2007 from http://www.census.gov/hhes/www/income/histinc/hstchg.html.

[41] Veracierto, M., (2008), *On the cyclical behavior of employment, unemployment and labor force participation*, Federal Reserve Bank of Chicago, February.

[42] Wachter, M.L., (1978), *Intermediate Swings in Labor Force Participation*, BPEA2, pp. 545– 576.

[43] West, K.K., Robinson, J.G., (1999), *What Do We Know About the Undercount of Children?* US Census Bureau working paper, Issued August 1999. http://www.census.gov/population/www/documentation/twps0039/twps0039.html

THE IMPACT OF ADVERTISING ON AGGREGATE CONSUMPTION: THE CASE OF ITALY

Luigi **MARATTIN**
University of Bologna, Department of Economics, **ITALY**
luigi.marattin@unibo.it

Abstract:
One of the last assumptions of neoclassical economics that has not yet been fully challenged is the exogeneity of consumers'preferences. In this paper we attempt to verify and measure the effects of advertising on consumers'demand.

We do so by carrying out an econometric analysis, relying on a rather simple econometric model on Italian economy, using quarterly data from 1980 to 2000. We build an ADL model with Koyck transformation and different (and advertising-specific) geometrical rates of decline, and we test both the flow-form and the stock form of advertising. Our conclusions show that in the period under consideration advertising had a positive and significant effect on consumption, with short term and long term elasticities equal respectively to 0.034 and 0.16.

Keywords: ADL models, advertising, Koyck transformation.

JEL classification: D12, D83

1. Introduction

In the continuous effort to update the enormous literature on consumers' demand, a particular feature has traditionally received very little attention: the investigation of the possible effects of advertising on aggregate consumption. The reason of this scarce recognition is to be found in the fact that neo-classical economies generally relies on a certain set of assumptions; among them, perfectly competitive economy with no uncertainty, market clearing and exogeneity of tastes and technologies. While virtually all of them have been relaxed within the increased realism and complexity of economic modelling, the last ones still stand as crucial milestone of economic theory.

The central issue of this paper is to challenge the exogeneity of preferences from a merely empirical point of view, by verifying and quantifying the impact of advertising on aggregate demand.

In Italy, the country this study is devoted to, very little research has been directed at any macro-implications of the effect of advertising, despite millions of euro being increasingly devoted to marketing and expenditure over recent years. Moreover, no italian study is known to have included a proxy for advertising expenditure among the explanatory variables in the aggregate consumption function.

The remainder of this paper is organized as follows. Section 2 offers a brief review of the existing empirical literature . Section 3 presents the econometric model, an Autoregressive Distributed Lags framework relying on Koyck-transformation tools; the two subsections explore the two alternative ways to consider advertising.. Section 4 addresses the cointegration and simultaneity issues and also analyses the short-run dynamic by building an Error Correction Model. Section 5 offers some concluding remarks.

2. Literature review

Empirical studies are found both to support and oppose a positive relationship between advertising and consumption; here we present a small survey of these opposite views, dividing them according to their final predictions.

Most contributions find a positive effect. Jung and Seldon (1995a) regress aggregate consumption on lagged consumption and advertising, finding a positive relationship. However, they only used a bivariate model, and this obviously weakens the argument, as advertising could capture the effects of a number of other variables which are usually though of as relevant in the specification of the consumption function (such as disposable income, interest rate and so on). In a second paper published in the same year, Jung and Seldon (1995b), they introduced cointegration techniques in their analysis on the relationship between advertising and consumption, based on U.S. data. They found that the two variables are integrated of order one and that are cointegrated; consequently, they examined an Error Correction Mechanism and found a two-way causality between advertising and consumption. They also put forward a rather doubtful assertion, according to which the original increase in consumption (triggered by advertising) could be offset by a decrease in investment,because of the decrease in savings, which in turn was caused by the increase in consumption; the overall impact on aggregate demand, therefore, would be nullified.

Keir (1993) used quarterly UK data from 1970 to 1991; he recognised the potential endogeneity of advertising and tested it using the Hausmann tests, thereby finding that the hypothesis of endogeneity cannot be rejected. Therefore, there is a dual relationship between consumption and advertising: the former causes the latter, and the latter causes the former. He deflates advertising expenditure with a television advertising price index to obtain the messages per 1000 homes. He calculated that in the long run a £1 increase in advertising leads to a £8.97 increase in consumption.

Peel (1975) used quarterly data for United Kingdom from 1956 to 1966. He estimated two different specification of the consumption function, a Keynesian one and a permanent income one, both augmented by advertising: in the first case, short run and long run impact are respectively found to be of £ 2.15 and £ 4.48, whereas in the second case of 0.05 and 0.06. He obtained similar results when he explicitly considered the issue of simultaneity, by using a system of equations. The overall conclusion is that advertising has a significant impact on aggregate consumption, although his approach does present some weak points, such as a non-completely accurate specification of the permanent income consumption function, and some doubts in the econometrics (for instance, he only tested for first-order autocorrelation although using quarterly data). It is interesting to note that Keir's estimate of the long-run impact of advertising is twice as much as the one found by Peel.

Brack and Cowling (1983) proposed a slight change of perspective; they tested empirically their theoretical view, according to which advertising affects consumption via the labour supply response. They argue that previous study on advertising's importance have usually focused on the propensity to consume, which has proved fairly constant over the long-run. However increasing the average propensity to consume is only one of a number of possible responses to advertising. According to them, the increase in desires caused by advertising can only be fulfilled by an increase in income for a large part of the labour force, since their average propensity to consume already tends towards one, and therefore their artificially increased needs encounter a budget constraint. Thus they investigated the possibility that advertising induces labour force to work longer hours than would be the case if advertising never took place. They use a measure of the advertising stock, rather than the flow, in order to be able to capture the cumulative effect. Following a growth of real advertising messages per head of 150%, a decay rate of 0.75 and an advertising elasticity of 0.18, the work year was approximately 27% longer than it would have been in 1976 in absence of advertising.

Taylor and Weiserbs's work (1972) was based on an extension of Houthakker-Taylor state adjustment model. They assumed that the effect of advertising on consumption is direct and that it operates through a flow rather than state variable; so if the flow suddenly stopped, the preference map (which is assumed to be modified as far as the marginal rate of substitution of consumption for saving is concerned) would revert to the shape it had before there was advertising. Their results suggest a positive impact of advertising on consumption; they found that a $1 per capita increase in advertising expenditure leads to an increase in per capita consumption of $4.55 in the short run and $7.85 in the long run. However, in evaluating their results, they highlight some of the aspects that might question the validity of the results. Particularly, they argue that the result may be spurious as advertising may act as a proxy for some other factor; furthermore, they find that the presence of simultaneity cannot be ruled out.

However, a number of studies contradicts the existence of a positive and significant effect of advertising on aggregate consumption. Schmalensee (1972) uses instrumental variables estimation and adds lagged, current and future advertising to the consumption function, finding that future advertising outperformed the current one, which in turn outperformed the past in fitting consumption data. Although no formal tests were applied, he used this result to imply that causation runs from consumption to advertising, and not the other way round.

In a later paper with Ashley and Granger (1980), he attempted to address the issue of causality more formally. They found that the post sample mean squared error of the model containing equations for advertising causing consumption as well as those for consumption causing advertising (the bivariate model) is 5.1% lower than the univariate model (containing only equation for consumption causing advertising). Therefore, they conclude that the bivariate model is not a significant improvement, and retain their null hypothesis that aggregate advertising does not have any role in explaining aggregate consumption. However, a major criticism of this paper is that they used raw consumption data along with seasonally adjusted quarterly advertising data, which is likely to introduce bias into the analysis of causality.

Pitelis (1987b) drew the attention to the effect of advertising on consumption through profits. According to him, the relationship works through two different ways: on one hand, increased profits lead to decreased consumption, given the assumption of a lower propensity to consume out of profit income than out of wages. On the other, more profits, lead also to increased retained profits, which will reduce consumer's expenditure, provided the retained profits are not perfectly substitutable for personal savings (a result proved in Pitelis (1987a)). Using quarterly data for United Kingdom from 1960 to 1972, he found that advertising does not cause consumption directly, but then he goes on to examine the relationship between advertising and profits: profits do not cause advertising, but advertising is highly significant in explaining profits. Therefore, while he casted doubts on any direct positive effect running from advertising to consumption, he found an indirect negative link through profits: advertising increases profits, which in turn decrease consumption.

To summarise, we have showed that different empirical studies (referring mainly to US and UK) are found both to support and oppose a positive relationship between advertising and aggregate consumption. Even among those studies that agree on a positive impact, there is quite a great variety in the magnitude of this effect.

3. The econometric model

We use quarterly data from 1980:1 to 2000:4. The dependent variable is real private consumption expenditure (source OECD database). Explanatory variables are GDP at constant prices (base year 1995, source OECD), real advertising expenditure deflated by the standard GDP deflator[15] (source Nielsen)[16], consumer price index (included to capture possible money illusion effects, source OECD) and past values of real private consumption (to capture habits and adjustment costs). Consumption function to be estimated is therefore:

$$C_t = f(Y_{t-i}, A_{t-i}, P_{t-i}, C_{t-i}) \qquad (1)$$

where: C – private consumption; Y – GDP; A – advertising; P – price index; with all variables being in logs.

But how is exactly advertising introduced into the model?

Earlier empirical analysis on the effect of advertising on consumption failed to recognize any dynamic effects, preferring to focus on the impact of current advertising on current consumption. Most of the recent economic studies have instead accepted that its impact is not fully dissipated in the period when that advertising takes place, and thus that any empirical model must include the effects of lagged values. This opinion seems reasonable for at least a couple of considerations. First, it may take a series of repeated advertising messages to break through a threshold of buying resistance; the last message, which actually triggers the purchasing, cannot be fully hold responsible for the success. Secondly, the potential consumer, once persuaded to buy the product, may not immediately purchase it; in other words, the consumer may well decide to purchase the good at time *t* (thereby marking the success of the advertising campaign) and actually going to the shop at time *t + 1*. Furthermore, we must take into account the possibility of lags between the investment in advertising by firms and the

[15]In response to a criticism againts the use of the GDP deflator to obtain a real measure of advertising expenditure, Jung and Seldon (1995b) found that the correlation between a specific cost-index for advertising and the GDP deflator is 0.992.

[16]Data show the following proportions of advertising expenditure shares: newspapers and magazines (59,47%), radio and televisions (37,54%),cinema (2%), other methods (0,78%).

moment where consumers actually get to see it: it is the case of advertising in durable media (e.g. magazines) that can be read months after the actual publication date, but also of Tv-advertising, which requires some technical time to produce the spot (hiring the director, actors, and so on).

Recognition of the dynamic effects of advertising has an immediate implication for the specification of the model; there are two alternative ways to acknowledge the dynamic role of advertising in its impact on aggregate consumption:

1) lagged values of advertising flow should be included as explanatory variables.

2) the string of advertising-flow variables should be combined so as to form a single variable representing advertising stock in the current period.

The following two subsections will investigate both options.

3.1 Advertising in flow form

In this formulation, the equation to be estimated is:

$$C_t = f(Y_{t-i}, P_{t-i}, A_{t-i}) \quad (2)$$

with i =[0,1,2,3……n]

Equation (2) is very likely to have an estimation problem posed by the presence of a long string of lagged regressors; degrees of freedom would disappear completely and even truncating the series several periods back, the remaining regressors would be likely to generate a high degree of multicollinearity in the model.

Here we assume that each of the three-lag structure in (2) takes a geometrically declining form, the so-called Koyck transformation.At this stage, we assume that the rate of decline λ is constant across the variables (we will abandon this assumption later on).

$$C_t = \alpha + \beta_1 P_t + \lambda\beta_1 P_{t-1} + ...\lambda\beta_1 P_{t-n} + \beta_2 Y_t + \lambda\beta_2 Y_{t-1} + ...\lambda\beta_2 Y_{t-n} + \beta_3 A_t + \lambda\beta_3 A_{t-1} + ...\lambda\beta_3 A_{t-n} \quad (3)$$

Rearranging we get:

$$C_t = (1-\lambda)\alpha + \beta_1 P_t + \beta_2 Y_t + \beta_3 A_t + \lambda C_{t-1} \quad (4)$$

with $0 < (1-\lambda) < 1$ being the rate of decline common to all the three lag structure. Equation (4) is the standard final form of the equation (to be taken to estimation) under the Koyck transformation; β_3 is the short-term marginal impact of current advertising on current consumption, whereas the long-term impact is given by $\frac{\beta_3}{(1-\lambda)}$.Equation (4) is a reasonable model, and rather convenient for estimation purposes. However, the assumption of constant rate of decline across variables seems too restrictive. Particularly, the lagged structure of advertising is likely to be determined by characteristics unique to that variable: for example, a series of advertising messages is supposed to patiently push an individual through a threshold of buying resistance, and this may require a different amount of time compared to what is needed for prices or income to have an effect on consumer demand. Therefore, we modify the model by allowing the rate of decline to differ across the variables. Technically, we assume that there are two kinds of dynamic effect in the model:

- the first common to all variables, each being given the same geometrically declining lag structure (rate of decline: $0 < (1-\varphi) < 1$).
- the second specific to advertising (rate of decline $0 < (1-\omega) < 1$)

Following this procedure, equation (4) is modified as follows:

$$C_t = (1-\omega)(1-\varphi)\alpha + \beta_1 P_t - \omega\beta_1 P_{t-1} + \beta_2 Y_t - \omega\beta_2 Y_{t-1} + (\omega+\varphi)C_{t-1} - \omega\varphi C_{t-2} + \beta_3 A_t \quad (5)$$

From equation (5) we work out our unrestricted model, where we can now distinguish the two above different dynamic effects:

$$C_t = \rho + \theta P_t + \tau P_{t-1} + \psi Y_t + \eta Y_{t-1} + \pi C_{t-1} + \delta C_{t-2} + \delta A_t \quad (6)$$

Rearranging (5):

$$C_t - (\omega + \varphi)C_{t-1} + \omega\varphi C_{t-2} = (1-\omega)(1-\varphi)\alpha + \beta_1(P_t - \omega P_{t-1}) + \beta_2(Y_t - \omega Y_{t-1}) + \beta_3 A_t$$

by renaming the coefficient according to:

$$C_t - (\omega + \varphi)C_{t-1} + \omega\varphi C_{t-2} = K_t$$

$$(1-\omega)(1-\varphi)\alpha = Z_t$$

$$P_t - \omega P_{t-1} = M_t$$

$$(Y_t - \omega Y_{t-1}) = \Phi_t$$

we obtain our restricted model:

$$K_t = \Phi_t + \beta_1 Z_t + \beta_2 M_t + \beta_3 A_t \quad (7)$$

In section 4 we will test the validity of the restrictions.

3.2. Advertising is stock form

The alternative to the flow-form is to combine the string of advertising-flow variables so as to form a single variable representing advertising stock in the current period. Defining this stock as:

$$G_t = A_t + \lambda A_{t-1} + \lambda A_{t-2} + ...\lambda A_{t-n} \quad (8)$$

we represent formally the alternative way to consider the impact of advertising on consumers' perception: the cumulative stock of advertising messages received until that moment is responsible for persuading potential buyers. Plugging (8) into the standard equation, we obtain:

$$C_t = \alpha + \beta_1 P_t + \beta_2 Y_t + \beta_3 G_t \quad (9)$$

which represents the stock-form model.

4. Cointegration and estimation

This section is concerned with the econometric analysis. First, issues of order of integration and simultaneity will be addresses, and subsequently we will estimate the unrestricted and restricted model presented in section 3. The last subsection builds an ECM to analyse the short term dynamics.

All variables are found to be I(1), according to the Dickey-Fuller procedure; Engle-Granger test show that they are also cointegrated, thereby pointing out a long run relationship among them.

The simultaneity issue is crucial. As argued in section 2, the empirical literature has long debated on the direction of causality between advertising and aggregate consumption. Not only, in fact, it is possible that advertising can cause consumption through the creation of wants, but it is also possible that consumption may cause advertising because higher consumption leads to higher profits, which in turn are the source of further funding for advertising. One of the assumption of Ordinary Least Squares regression is that explanatory variables are either non-stochastic or, if stochastic, they must be distributed independently of the stochastic disturbance term. If they are not, the OLS estimators are biased and inconsistent. Therefore it is vital to examine the question of potential endogeneity of advertising; we do so by applying the Hausmann test. First we postulate the consumer demand equation and the advertising equation, here in implicit form:

$$C_t = f(P_t, P_{t-1}, Y_t, Y_{t-1}, C_t, C_{t-1}, A_t) \quad (10)$$

$$A_t = f(A_{t-1}, C_t, C_{t-1}) \quad (11)$$

Then we regress advertising on the exogenous variables of (10) and (11) :

$$A_t = f(P_t, P_{t-1}, Y_t, Y_{t-1}, w) \quad (12)$$

with w being the stochastic residuals, to be included in the following equation for the final stage of the test:

$$C_t = const + \beta_1 P_t + \beta_2 P_{t-1} + \beta_3 Y_t + \beta_4 Y_{t-1} + \beta_5 w_t + u_t \quad (13)$$

with u_t being the error term.

According to the Hausmann-test procedure, if the coefficient β_5 is significant, there is a simultaneity problem, otherwise there is not.

Regression output shows a value for β_5 equal to 0.04491 , with standard error 0.036 and a corresponding p-value of 0.1289; therefore it is found not significant at 5% level. We conclude that for the time period under consideration and for the data available, there was not simultaneity between advertising and consumption in Italy.

We can now proceed with the estimation of the unrestricted model, equation (6) of the previous section. Table 1 reports the OLS estimation output:

Table 1. Estimation of unrestricted model

Variable	Coefficient	Standard Error
P_t	- 0,2272	0.0968
P_{t1}	0.1973	0.0935
Y_t	0.2841	0.0842
Y_{t1}	- 0,1113	0.0908
C_{t1}	0.2546	0.0989
C_{t2}	- 0,4333	0.0854
A	0.034	0.0169

As it can be seen from estimation output, the model presents rather satisfying results: all coefficients (apart from the lagged value of income) are significant at 5% level.

As far as elasticities are concerned, these are the results:

In the short run, price elasticity is estimated at 0.22, whereas income elasticity at 0.28. However, for the purpose of this study, the most relevant is the elasticity of consumption with respect to advertising, which is estimated at 0.034. Long run elasticity of advertising is estimated at 0.16.Therefore the long-term impact of advertising on consumption is more than four times greater than the short-term impact. The analysis of residuals shows no sign of autocorrelation; this hypothesis is confirmed by the analysis of correlogram and the Breusch-Godfrey Serial Correlation LM test.

Turning to the restricted model, prior to any estimation it is necessary to calibrate the values for the rates of decline $(1-\varphi)$ and $(1-\omega)$. Several simulations were carried out, and the best results were obtained with the values of, respectively, 0.03 and 0.05. They correspond to a rate of decline of 95% for advertising and 97% for other variables. This might be considered reasonable since it can be argued that past advertising has a slight greater effect on current propensity to consume than past values of income or price. Furthermore, this particular rate of decline of advertising has already been used in an empirical analysis on the topic (Malcom,1997).

Table 2. reports the estimation output for the restricted model:

Table 2. Estimation of the restricted model

Variable	Coefficient	Standard Error
Z	- 0,1656	0.4314
M	0.8370	0.085
A	0.070	0.40

where, as recalled in the previous section:

$$(1-\omega)(1-\varphi)\alpha = Z_t$$

$P_t - \omega P_{t-1} = M_t$

As it can be seen, the coefficient on the variable Z is negative just as well as it was the one on P in the unrestricted model, thereby confirming the validity of our construction. At this stage, we test the validity of the restrictions imposed on the model by using a F-test. The restrictions are as follows:

$\theta = \beta_1$
$\tau = -\omega\beta_1$
$\psi = \beta_2$
$\eta = -\omega\beta_2$
$\pi = \varphi + \omega$
$\delta = -\varphi\omega$
$\sigma = \beta_3$

F-test is constructed according to the usual:

$$F = \frac{\left(RSS_r - RSS_{ur}\right)/m}{RSS_{ur}/(n-k)} \tag{14}$$

where:
RSS_r =residual sum squares of restricted model
RSS_{ur} = residual sum squares of unrestricted model
m= number of linear restrictions
n= number of observations
k= number of parameters in the unrestricted model

In our case, the null hypothesis (the validity of the restrictions) is accepted at 5% significance level, as the application of (14) returns a value of 1.819 , to be compared with the following critical values:

$F_{0.01} = 2.82$
$F_{0.05} = 2.10$
$F_{0.10} = 1.77$

Thus, our set of restrictions seems to be valid.

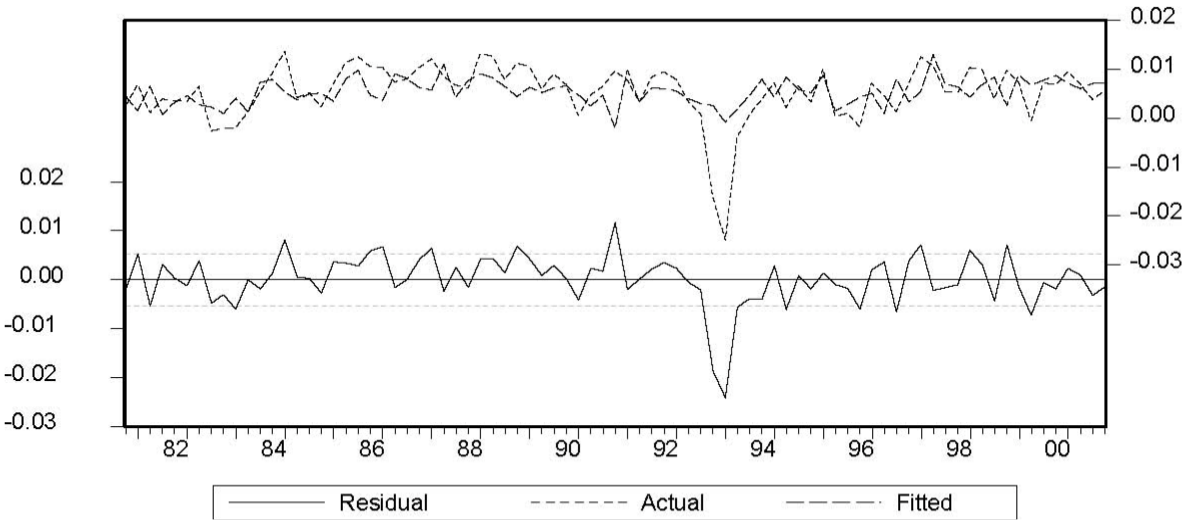
Turning our attention to the restricted model, Engle-Granger test shows that variables are I(0) and cointegrated. In order to describe more accurately the short-term dynamic of the restricted model, we build an Error Correction Model; since it has to include only stationary variables, we will use the first-difference of the previous series (indicated with a d before the corresponding letter). ECM includes also the residuals from the long-run equation (in this case, the restricted model), lagged by one period (indicated by RESID) Estimation output can be found in table 3, followed by the residual graph in Figure 1.

Table 3. Estimation of ECM

Variable	Coefficient	Standard error
dZ	-0,003	0.001

dM	0.38	0.009
dA	0.009	0.006
RESID	-0,2420	0.006

Figure 1: Residual graph



As it can be seen, 24.2 per cent of the deviation from the equilibrium error is corrected in each period.

Finally, table 4 reports the estimation output for the model with advertising in stock form (equation 9):

Table 4. Estimation of the stock-form model

Variable	Coefficient	Standard Error
P_t	-0,1873	0.028
Y_t	0.8709	0.32
G_t	0.075	0.417

The coefficient on the stock variable G is significant, with elasticity equal to 0.075. Previously, by considering advertising in flow-form, we found an elasticity of 0.034.

Thus, recognizing the dynamic effect of advertising by considering it as flow-form or stock form does not change either qualitatively nor quantitatively the main conclusion of this paper: advertising has a positive impact on aggregate consumption, whose magnitude can reasonably be estimated between elasticity of 0.034 and 0.075.[17]

5. Conclusions

This paper attempted to analyse and measure the effects of advertising on aggregate consumption dynamics in Italy from 1980 to 2000. The empirical analysis has been conducted by building a model with different and exogenously-determined rates of decline for the explanatory variables of aggregate consumption (the dependent variable), among which aggregate advertising expenditure has been included for the first time in the national economic literature. The results can be summarized as follows. There seems to be a positive and significant impact of advertising on consumption; specifically, short run and long run elasticities are found to be, respectively, 0.034 and 0.16; we have also shown that the results are approximately the same if we consider advertising in stock form instead of flow form. The nature of the rather simple technical analysis leaves plenty of

[17]We also estimated the stock form model with the inclusion of lagged values of prices and income, and also in that case the advertising elasticity approximates the value reported in the paper.

room to improve the adequacy of the investigation; the biggest obstacle in that sense is no doubt the lack of an appropriate dataset, able to account for cross-country differences and allowing researchers to better identify and measure the way advertising expenditure has affected consumption levels in our economies.

6. References

[1] Ashley, R., Granger, C.W.J., Schmalensee, R., (1980), *Advertising and Aggregate Consumption: An Analysis of Causality*, in: *Econometrica*.

[2] Brack, J., Cowling K., (1983), *Advertising and Labor Supply: Workweek and Workyear in U.S. Manufacturing Industries*, 1919-76. Kyklos - *International Review for Social Studies*.

[3] Hausmann, J.A., (1976), *Specification Tests in Econometrics*, in: *Econometrica*, vol.46, November 1976, pp.1251 – 1271.

[4] Jung, C., Seldon, B.J., (1995a), *The lenght of the effect of aggregate advertising on aggregate consumption*, in: *Economic Letters*.

[5] Jung, C., Seldon, B.J., (1995b), *The Macroeconomic Relationship Between Advertising and Consumption, in: Southern Economic Journal*.

[6] Keir, T., (1993), *The Aggregate Advertising- Consumption Relationship*, Revisited mimeo – University of Warwick

[7] Peel, D., (1975), *Advertising and Aggregate Consumption in Advertising and Economic Behaviour* ed. K.G.Cowling, J.C.Cable, M.K.Kelly and T. (1975) McGuinness (McMillan Press Ltd.).

[8] Pitelis, C.N., (1987a), *Corporate Capital*, Cambridge University Press.

[9] Pitelis, C.N., (1987b), *The Causal Relationship Between Advertising, Retained Profits and Aggregate Consumption. A Note*, in: *Greek Economic Review*.

[10] Schmalensee, R., (1972), *The Economic of Advertising*, North Holland Publishing Company.

[11] Taylor, Weiserbs, (1972), *Advertising and the Aggregate Consumption Function,* in: *American Economic Review*.

THE ECONOMIC CAPITAL OF OPAQUE FINANCIAL INSTITUTIONS

Fernando **MIERZEJEWSKI**
Faculty of Business and Economics
Katholieke Universiteit Leuven, **BELGIUM**
Fernando.Mierzejewski@wis.kuleuven.be, femierze@lycos.com

Abstract:
The capital structure of firms that cannot hedge continuously is affected by the agency costs and the moral-hazard implicit in the contracts they establish with stockholders and customers. It is demonstrated in this paper that then an optimal level of capital exists, which is characterised in terms of the actuarial prices of the involved agreements. The capital principle so obtained extends the classic theoretical framework, sustained by the well-known proposition of Modigliani and Miller and the model of deposit insurance of Robert Merton, at the time that naturally integrates the financial and actuarial theoretical settings.

Keywords: capital structure, economic capital, risk capital, deposit insurance, value-at-risk.

JEL classification: G11, G30, G31, O16

1. Introduction

A model will be presented to characterise the level of capital demanded by firms that access to capital markets where hedging is restricted to some extent.

At the empirical level, the implied capital principle shows advantages over other principles found in the literature. Firstly, it is founded on economic fundamentals. Secondly, since it is expressed in terms of the quantile function of the underlying risk, it can be applied to any kind of probability distributions and hence, it is suitable both to finance and insurance applications. The quantile function is actually well-known by researchers and practitioners and it has been recommended by the Basel Committee (2004).

2. Economic Capital in Opaque Institutions

Financial institutions maintain cash provisions, in the form of capital, to guarantee that their contracted liabilities will be honoured at the end of the investment period.

Several types of capital are distinguished in the literature. Thus, on the one hand, cash capital represents a balance required to execute transactions, while working capital additionally includes operational expenses [see Williams *et al*., (2002)]. On the other hand, the term regulatory capital refers to a balance defined according to some accounting standard [as in Basel, (2004)], while equity corresponds to the portion of reserves provided by shareholders. Finally, many authors speak of economic or risk capital when some particular criterion, based on economic or statistical considerations, is proposed.

In the lines with Merton (1974, 1977), the demand for capital will be corresponded to a demand for deposit insurance in this paper. Accordingly, the terms economic and risk capital will be indistinctly used to refer to the smallest amount required to insure the value of the portfolio of net assets $X = A - D$, where A and D respectively denote the market values of the portfolios of outstanding assets and liabilities.

In this context, the difference between economic and equity capitals represents a balance that is supplied by managers in attention to some solvency requirement.

Three main components of the capital structure can then be distinguished: a net liability contracted with customers, an amount of equity supplied by shareholders and a cash balance provided by managers. Since the economic capital is equal to the sum of the last two components, the problem of capital allocation can be roughly corresponded to the determination of the proportions of the portfolio of assets that are funded by means of internal and external debt.

Holding capital imposes an opportunity cost on firms because these funds could be alternatively employed on profitable investments. Such costs lead financial institutions to prefer external debt and accordingly demand as less capital as possible.

In fact, in a seminal paper, Modigliani and Miller (1958) claim that, if at any moment firms can borrow and lend any amount of capital at a single interest rate, they can adjust their balance sheets whenever is needed and hence cash provisions impose a cost without any benefit. Then rational decision-makers (who maximise value) should demand no capital at all.

More specifically, Merton (1997) states that the presence of credit-sensitive customers obliges opaque institutions (whose investment activities are not fully observed by outsiders) to rely on a third-party guarantor, who agrees to honour the outstanding liabilities when bankruptcy is declared. The presence of credit-sensitive customers thereby increases external controls and monitoring due to the moral-hazard implicit in the administration of deposits [see also Ross, (1989)].

In this context, the market values of equity and debt can be respectively corresponded to the values of a call and a put option on the value of assets, with exercise price equal to the value of debt, which implies that the market value of the firm (or the market value of the assets' portfolio) is independent of the capital structure, as predicted by the MM-proposition [see also Cummins and Sommer, (1996), and Miller, (1998)]. A fundamental assumption for this mechanism to work is that capital and financial securities are continuously traded in competitive markets.

The correspondence of capital to deposit insurance implies that the hypothesis of perfect hedging can be formulated by imposing that every claim is assigned a unique insurance price in the market. Under such circumstances, managers are indifferent between hedging and insurance and are certainly indifferent about the amount of economic capital.

However, hedging and insurance cannot be regarded as equivalent tools in practice. As a matter of fact, although competitive forces in capital and security markets may lead transactions to be produced at a unique price, this is scarcely the case of insurance markets, where non-standardised policies are transacted [see Goovaerts *et al*., (2005), and Venter, (1991)].

Liquidity restrictions may thereby arise from two different sources. In the first place, firms are not always able to trade continuously in capital and security markets. Then the moral-hazard arising due to the opaqueness of financial intermediaries induces the appearance of premiums over the market cost of capital, which should be established on an actuarial basis. Secondly, the buyers and sellers of insurance can maintain different perceptions about risks and can accordingly assign different prices to their corresponding guarantees.

The optimal cash balance will be determined in this paper, in such a way that the market value of the firm, defined as the difference between the actuarial prices of equity and the default claim, plus the return offered by a non-risky bond, is maximised.

A precise description of the conditions under which the capital structure matters will be provided in this way. Hence the limits of the Modigliani and Miller invariance propositions will be clearly stated.

3. The Optimal Capital Structure

Both asset and liability claims will be regarded as random variables in the following, while the market value of net assets will be expressed as the product of the level of investment I and some random perturbations X:

$$A - D = I \cdot X \quad \Rightarrow \quad X = \frac{A - D}{I} \tag{3.1}$$

Then the level of investment can be regarded as the principal of the net portfolio. The level of capital will be also represented as a proportion of the level of investment:

$$K = I \cdot k \quad \Rightarrow \quad k = \frac{K}{I} \tag{3.2}$$

The ratio k represents a capital-to-investment or a cash-to-risk ratio. It determines the proportion of internal financing of the firm.

Every contract of deposit insurance obliges the guarantor to honour the total capital loss:

$$I \cdot (X + k)_{-} = I \cdot \min\{0, X + k\} \tag{3.3}$$

in exchange of a certain premium paid by the former.

Simultaneously, shareholders pay a certain price to managers at the beginning of the investment period, in exchange of receiving the following random capital profit at the end:

$$I \cdot (X-k)_+ = I \cdot \max\{0, X-k\} \tag{3.4}$$

The capital K is invested in a banking account to obtain the risk-free return r_0 (or in other words, it is converted to a risk-free zero-coupon bond with internal return r_0, promising to pay the cash flow $r_0 \cdot K$ at some maturity date) [see Hull, (2000)].

The payments of the net portfolio at maturity are then given by:

$$\begin{aligned} &I \cdot (X-k) + r_0 \cdot K \quad && if\ X \ge k \\ &I \cdot (X+k) + r_0 \cdot K \quad && if\ X \le -k \\ &0 \quad && if\ -k < X < k \end{aligned} \tag{3.5}$$

Hence the firm can afford its debt and pay a surplus to shareholders when $X \ge k$, i.e. the firm is solvent in this case.

By contrast, when $X \le -k$ the capital K is deliver to the guarantor who has to afford the residual loss $I \cdot (X+k)$. Shareholders receive nothing in this case.

Finally, when $-k < X < k$ the firm cannot return the total amount of capital to shareholders, although the total debt attracted from customers can be honoured and the guaranty is not invoked. Stockholders might decide to sell their shares or to call for portfolio restructuring under such circumstances.

We have already pointed out that risk can be completely suppressed through hedging if cash and securities can be traded to any desired extent in the market.

Indeed, under such circumstances, the prices (per unit of investment) of the contracts established with shareholders and guarantors are respectively given by the prices of a call and a put option on the value of the random capital return X with exercise price equal to the cash-to-risk ratio k, in such a way that the put-call parity can be invoked to obtain:

$$X = C(X,k) + k \cdot e^{-r_0 \cdot T} - P(X,k) \tag{3.6}$$

where T denotes the time to maturity. Then the value of the firm does not depend on the cash-to-risk ratio k, as established in the MM-proposition, as long as continuous rebalancing of portfolios is allowed [Merton, (1977), Cummins and Sommer, (1996), Miller, (1998)].

When hedging cannot be implemented to any extent due to liquidity restrictions, the prices of the contracts established with stockholders and guarantors should be determined on an actuarial basis.

Accordingly, the price of equity and the price of insurance (equal to the cost of bankruptcy) must be respectively corresponded to the following terms:

$$\begin{aligned} E[(X-k)_+] &= \int_{+k}^{+\infty} (x-k) \cdot dF_X(x) \\ &= \int_{+k}^{+\infty} (x-k) \cdot f_X(x) dx \end{aligned} \tag{3.7}$$

$$\begin{aligned} E[(X+k)_-] &= -\int_{-\infty}^{-k} (x+k) \cdot dF_X(x) \\ &= -\int_{-\infty}^{-k} (x+k) \cdot f_X(x) dx \end{aligned} \tag{3.8}$$

for these terms represent the actuarial prices of the corresponding underlying exposures [see Goovaerts *et al.*, (1984)].

In Eq. (3.7) and (3.8), the terms F_X and f_X respectively denote the cumulative and density probability distributions, respectively defined as [see, for example, De Finetti, (1975)]:

$$F_X(x) = P\{X \le x\} \tag{3.9}$$

$$f_X(x) = P\{X = x\} = \frac{dF_X(x)}{dx} \tag{3.10}$$

Then the value of the firm at the end of the investment period (as perceived by managers) is given by:

$$V = \left(E[(X-k)_+] - E[(X+k)_-] \right) \cdot I + r_0 \cdot K \tag{3.11}$$

Within this context, the cash-to-risk ratio k affects the net return on investment and hence the value of the firm.

Given any fixed level of investment I, every rational manager must choose the capital structure that maximises the firm's value per unit of investment V/I. From Eqs. (3.1), (3.2), (3.3), (3.4), (3.7), (3.8) and (3.11) we obtain that every rational manager must then solve the following maximisation problem:

$$\max_k E[(X-k)_+] - E[(X+k)_-] + r_0 \cdot k \tag{3.12}$$

An alternative model that characterises the optimal capital structure in terms of the values $E[(X-k)_+]$ and $E[(X+k)_-]$ is suggested by Froot *et al.*, (1993). In that model, the market value of the portfolio depends on some absolute perturbation X, and not a relative perturbation, as it is the case in the model presented in this paper, see Eq. (3.1).

The solution to the maximisation problem (3.12) can be determined by applying Lagrange optimisation. The first-order condition actually leads to:

$$\frac{dE[(X-k^*)_+]}{dk} - \frac{dE[(X+k^*)_-]}{dk} + r_0 = 0 \tag{3.13}$$

As noticed from Eqs. (3.7) and (3.8), the derivation of the terms $E[(X-k)_+]$ and $E[(X+k)_-]$ implies the derivation of integral operators with respect to a variable that affects the limits of integration. The Leibnitz's rule can then be applied:

$$\begin{aligned} \frac{d}{dz}\int_{u(z)}^{v(z)} \varphi(z,x)dx &= \int_{u(z)}^{v(z)} \frac{\partial \varphi(z,x)}{dz}dx \\ &+ \varphi(z,v(z)) \cdot \frac{dv(z)}{dz} - \varphi(z,u(z)) \cdot \frac{du(z)}{dz} \end{aligned} \tag{3.14}$$

From Eqs. (3.7) and (3.14) we obtain that:

$$\begin{aligned} \frac{dE[(X-k)_+]}{dk} &= -\int_{+k}^{+\infty} dF_X(x) \\ &= -P\{X > k\} = -T_X(k) \quad \forall k \end{aligned} \tag{3.15}$$

where the function $T_X = 1 - F_X$ denotes the tail or survival probability distribution.

Besides, since $P\{X \le -k\} = P\{-X > k\}$, we obtain from Eqs. (3.8) and (3.14) that:

$$\begin{aligned} \frac{dE[(X+k)_-]}{dk} &= -\int_{-\infty}^{-k} dF_X(x) \\ &= -P\{-X > k\} = -T_{-X}(k) \quad \forall k \end{aligned} \tag{3.16}$$

From Eqs. (3.13), (3.15) and (3.16) we finally arrive to the following expression for the first-order condition of the maximisation problem of Eq. (3.12):

$$T_{-X}(k^*) + r_0 = T_X(k^*) \tag{3.17}$$

which can be equivalently established in terms of the cumulative probability functions:

$$1 - F_{-X}(k^*) + r_0 = 1 - F_X(k^*) \tag{3.18}$$

Therefore, the optimal cash balance is determined at the point where the marginal reduction in the excess of cash over the level of capital, determined in Eq. (3.15), equals the marginal return on investment, represented by the term of Eq. (3.16), plus the risk-free interest rate r_0.

The existence of a solution to the maximisation problem (3.12) can be mathematically assured as long as the related objective function is concave, or in other words, as long as the second derivative of the objective function is lower than zero, i.e. as long as:

$$\frac{d^2E\left[\left(X-k^*\right)_+\right]}{dk^2}-\frac{d^2E\left[\left(X+k^*\right)_-\right]}{dk^2}<0 \tag{3.19}$$

From Eqs. (3.15) and (3.16), this condition is equivalent to:

$$\frac{dT_{-X}\left(k^*\right)}{dk}<\frac{dT_X\left(k^*\right)}{dk} \tag{3.20}$$

which from Eq. (3.10) and since $T_X = 1 - F_X$, can be still rewritten as:

$$P\left\{X=k^*\right\}<P\left\{X=-k^*\right\} \tag{3.21}$$

We thus arrive to the (reasonable) conclusion that capital is beneficial to financial institutions only when the probability of attaining a capital loss of a certain magnitude is greater than the probability of obtaining a capital gain of the same magnitude.

In particular, every symmetrical probability distribution (around its expected value) satisfies:

$$P\left\{X=E\left[X\right]-x\right\}=P\left\{X=E\left[X\right]+x\right\} \quad \forall x \tag{3.22}$$

Hence, within the class of risks characterised by symmetrical probability distributions, the objective function of the maximisation problem of Eq. (3.12) is concave when $E\left[X\right]<0$, but convex when $E\left[X\right]>0$. Accordingly, a maximum exists in this case only when $E\left[X\right]<0$. In other words, capital is beneficial for this kind of financial institutions only when $E\left[X\right]<0$.

4. The Cost of Capital of Opaque Intermediaries

As demonstrated in the previous section, in the presence of liquidity restrictions, the market value of financial institutions, as determined by Eq. (3.11), may well be affected by the proportion of internal finance k ─ and hence by the underlying capital structure.

Liquidity restrictions arise, in the first place, because the portion of capital provided by stockholders is determined in a regular frequency and cannot be modified until the end of the investment period. As a matter of fact, the amount of equity and the frequency of revisions are the result of negotiations between managers and stockholders. Changing such agreements is necessarily costly − and can reduce the market valorisation of the firm.

On the other hand, choosing regulatory and risk-based capital principles implies that the amount of economic capital must be subject to constant revisions, as long as the risk ness of the net-assets' portfolio is varying − i.e. as long as the series of capital profits and losses of the net-assets' portfolio is non-stationary. This means that managers are obliged to rely on some market of cash balances (or inter-bank loans) in order to maintain a total level of capital that is consistent with the underlying exposure.

Although preferring external debt reduces the controls imposed by shareholders, this strategy also raises the costs associated to moral-hazard (on the side of customers) and bankruptcy. This is especially true in highly leveraged firms, where managers have strong incentives to take risk [see Jensen, (1986)].

Consequently, when deciding their capital structures, firms have to face a trade off between paying high spreads because of opaqueness and signalling costs on the one hand, and sacrificing potential competitive advantages when maintaining idle balances on the other [see also Fama, (1980), and Ross, (1989)].

Another kind of liquidity restrictions thereby arises due to the fact that the opportunity cost of capital, that is perceived by managers as the reduction in the price of equity induced when certain level of capital is maintained, is not necessarily equal to the return r they have to pay to borrow in the market of cash balances, since in general:

$$-\frac{dE\left[\left(X-k^*\right)_+\right]}{dk}=T_X\left(k^*\right)\neq r \tag{4.1}$$

Indeed, while the cost of equity reflects the agency costs between managers and stockholders, the market capital cost r is determined according to the capacity and willingness to pay of the borrower and it then reflects the moral-hazard in the relationship with customers. It depends, in particular, on the capital structure of the borrower institution.

When determining the price of loans, the creditors of opaque organisations rely on their own research and monitoring, as well as on the information published in the media and the risk categorisations provided by specialised (private and governmental) institutions.

The credit ratings observed in practice normally include a finite number of categories. Within each class, every firm is supposed to face the same risk of default and hence every firm is allowed to borrow at the same interest rate, in such a way that the more the concerns of creditors about the credit capacity of firms in a certain class, the higher the level of the corresponding cost of capital and vice-versa.

This means that lenders cannot discriminate perfectly and that borrowers can remain in the same class as long as they do not drastically modify their capital structures. In other words, as long as firms do not drastically vary their cash-to-risk ratios, they can regard the market capital cost as a constant.

Let us consider in the following some firm that belongs to a certain class determined by the capital cost r and maintains the cash-to-risk ratio k.

The cost of equity of this kind of firms is given by:

$$E[(X-k)_+] = E[X_+] - r \cdot k \tag{4.2}$$

This means that the external and internal financing alternatives of the firm, respectively characterised by the return r and the difference in the price of equity when including the stock of capital k, are at equilibrium only when:

$$\frac{E[X_+] - E[(X-k)_+]}{k} = r \tag{4.3}$$

A capital profit and loss is then expected to be produced, due to differences between the internal and external estimations of the cost of capital, since in general we expect that:

$$\Delta = \frac{E[X_+] - E[(X-k)_+]}{k} - r \neq 0 \tag{4.4}$$

Accordingly, external funding is regarded as expensive by those firms that obtain $\Delta < 0$, for in this case maintaining the cash-to-risk ratio k produces a loss that is lower than the alternative cost of borrowing the same balance in the market. These firms prefer to demand reserves instead of relying on external finance.

Conversely, external finance is cheap for those firms that obtain $\Delta > 0$, for they have to incur in a higher loss when they choose to maintain a stock of capital. These firms prefer to demand no capital at all.

Jensen (1986) has noticed that internal monitoring is more intense when positive balances are obtained at the end of the investment period and cash is at disposal in excess of what is required to fund every ongoing (solvent) investment project. In this case, which can be actually corresponded to the condition $\Delta > 0$, it is said that a firm owns free-cash-flow.

In the model of Jensen (1986), frictions and mismanagement are specially severe within firms disposing of high amounts of free-cash-flow, resulting from the competition, between managers and shareholders, to take control of the profits generated by the company.

Since competition in product and factor markets should push utilities to a minimum level (eventually to zero), only those activities generating substantial economic rents are supposed to generate substantial amounts of free cash flow. Such activities are corresponded to product and factor markets where market forces are weak and where monitoring is more important than ever.

In conclusion, short-term stickiness inherent in the equity contracts established with stockholders, as well as in the credit categorisations determined by lenders in the markets of cash balances, may prevent financial institutions from continuously adjusting their capital structures.

Thus, on the one hand, changing the amount of equity implies a redistribution of the cash flow at disposal inside the institution that may increase the agency costs implicit in the relationship between managers and stockholders (as stated by Jensen) and which can be explicitly measured in terms of the actuarial price of the excess claim $X-k$.

On the other hand, raising the amount of external debt may raise the bankruptcy costs faced by institutions and the moral-hazard inherent in their relationship with creditors. These adjustments are performed through adjustments in the cost of capital r ─ or in other words, in the risk categorisations of the lender institutions.

5. Economic Capital as the Optimal Deductible

An optimal capital principle can be derived, based on an optimal compromise of bankruptcy costs and the market price of external debt.

Indeed, first notice that replacing Eq. (4.2) in Eq. (3.11) we obtain that the value, per unit of investment, of those firms that can borrow at the interest rate r is equal to:

$$\frac{V}{I} = E[X_+] - E[(X+k)_-] - (r - r_0)\cdot k \qquad (5.1)$$

Maximising value is then equivalent to minimise the total burden of default, equal to the price of insurance, represented by the term $E[(X+k)_-]$, plus the net benefit obtained when investing capital at the interest rate r instead of maintaining it at the low (non-risky) rate r_0:

$$\min_k E[(X+k)_-] + (r - r_0)\cdot k \qquad (5.2)$$

This problem has been already used to derive a rule of capital allocation by Dhaene *et al.* (2003), Laeven and Goovaerts (2004) and also Goovaerts *et al.* (2005). They regard its solution as an optimal solvency margin, which establishes a compromise between the costs of capital on the one hand, and a solvency requirement on the other. When justifying the implementation of this rule, they emphasise that arbitrage opportunities are difficult to exploit in insurance markets.

The first-order condition of the minimisation problem (5.2) leads to:

$$\frac{dE[(X+k^*)_-]}{dk} + r - r_0 = 0 \qquad (5.3)$$

which from Eq. (3.16) can be rewritten as:

$$-T_{-X}(k^*) + r - r_0 = 0 \qquad (5.4)$$

Besides, since:

$$\frac{d^2E[(X+k)_-]}{dk^2} = \frac{dF_{-X}(k)}{dk} = f_X(k)$$
$$= P\{X = -k\} > 0 \quad \forall\, k$$

(5.5)

we obtain that the objective function of the minimisation problem (5.2) is always convex, and hence, this function always attains a minimum.

More specifically, as long as the marginal benefit of adding the first unit of capital is greater than the net investment premium, i.e. as long as:

$$T_{-X}(0) > r - r_0 \qquad (5.6)$$

and as long as some capital loss is produced with non-zero probability, a range exists where the term $E[(X+k)_-]$ is convex in k. In this case, a level of capital exists that minimises the criterion of Eq. (5.2).

Under such circumstances, the optimal capital demand is determined by the quantile function of the probability distribution of the series of capital profit and losses of the underlying risk:

$$k^* = T_{-X}^{-1}(r - r_0) = F_{-X}^{-1}(1 - r + r_0) \qquad (5.7)$$

or equivalently, in terms of the liquidity premium $\nu = r - r_0$:

$$k^* = T_{-X}^{-1}(\nu) = F_{-X}^{-1}(1 - \nu) \qquad (5.8)$$

In other words, the optimal surplus is expressed as the Value-at-Risk (or VaR) for the confidence probability level $\nu = r - r_0$ (the definition and interpretation of the Value-at-Risk can be found in [Hull, (2000)]).

The VaR has been recommended for the implementation of good risk management practices by the Basel Committee on Banking Supervision (2004).

The fact that the confidence level in the definition of VaR is replaced by a net premium in Eq. (5.7) is a consequence of the first-order condition, which determines an exchange between a sure flow and a flow of probability ─ as specified in Eqs. (3.17), (3.18) and (5.4).

Accordingly, the higher the liquidity premium ν (i.e. the more the free-cash-flow at disposal), the more expensive to maintain a cash balance and hence the less the capital demanded. Conversely, the lower the liquidity premium, the cheaper the capital and hence the more the quantity demanded of this resource. The minimum and the maximum levels are respectively chosen when $\nu \geq 1$ and when $\nu \leq 0$.

From the actuarial viewpoint, the expected excess loss $E[(X+k)_-]$ represents the fair price of a special insuring contract (sometimes called layer) that obliges the insurer to pay to the policyholder the excess of loss over the level k, when such a loss is produced [see Goovaerts, (1984)].

In this context, the amount k represents a guarantee provided by the policyholder in order to assure the insurer (up to some extent) that every reasonable care will be taken to reduce the underlying exposure. In other words, the guarantee k, which is known as the deductible or retention in the literature, is introduced in insurance contracts as a means of reducing the costs derived from moral-hazard.

Within this framework, the optimal level of capital corresponds to the optimal deductible or the optimal retention of the related insuring liability contract.

Notice that full-coverage is implicitly assumed in the model, because the actuarial prices of equity and insurance have been expressed in terms of mathematical expectations that consider unlimited losses and gains over the level of capital. However, insurance contracts always specify a maximum payment in practice. The question then arises of who does eventually bear the risk of deposits.

According to the terms of the guaranteeing contracts previously defined, we can say that risk-bearing is roughly distributed in the following way: any loss up to the retention level k is paid by the firm (recall that shareholders only endure the equity component of the economic capital); losses that are higher than the retention level are paid by the guarantor or insurer, as long as these losses do not surpasses a maximum disaster level M^{DIS}; finally, some companies can seek for additional protection by establishing a contract with some reinsurance institution that agrees to pay any loss greater than the disaster level M^{DIS}, but lower than some catastrophe level M^{CAT}.

Hence, in the case of catastrophic events, it is the society as a whole who has to afford the losses ─ through governmental divisions, private creditors, companies and householders. This explains why it is in the interest of regulators to define good practices and regulatory requirements that can induce financial intermediaries to seek for protection according to the risk they actually bear.

From the economic point of view, the existence of an optimal level of capital implies that choosing a different level necessarily leads to over- or under-investment.

Indeed, idle money (which could be used to fund profitable investments) is maintained in excess when more capital than the optimal level is demanded. By contrast, when the stock of capital is lower than the optimal level, risk is taken in excess, a fact that might eventually increase the frequency of losses (as well as disaster and catastrophic events) and induce investors to raise their concerns about the credit quality of the firm. The price at which the firm can attract debt in the market might increase under such circumstances.

Therefore, independently of whether managers consider or not any of the optimisation problems presented in this paper, their capital preferences should approach to the solutions of these problems, for only following this strategy the market value of firms can be maximised (or the burden of bankruptcy can be minimised).

On these grounds, the optimal capital principle provides a basis for the determination of the aggregate behaviour of markets or multidivisional corporations (I analyse the problem of the allocation of capital inside multibusinesses corporations in Mierzejewski, 2006).

6. The Optimal Gaussian Capital Principle

When the underlying exposure X follows a Gaussian probability distribution with mean μ and volatility σ the optimal capital principle, as defined in Equation (5.8), takes the form:

$$k(\nu) = \sigma \cdot \Phi^{-1}(1-\nu) - \mu \tag{6.1}$$

where ν denotes the liquidity premium and Φ denotes the cumulative probability function of a standard Gaussian probability distribution, whose mean and volatility are respectively equal to zero and one, and which is defined as (see De Finetti, 1975):

$$\Phi(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^{x} \exp\left(\frac{-y^2}{2}\right) dy \quad \forall x$$

(6.2)

Then the optimal Gaussian capital principle follows a straight line in the plane of cash-to-risk ratios and volatilities see Figure (6.1).

Also the option-based capital rule of Robert Merton shows a linear dependence on volatility.

Indeed, as stated by Merton and Perold (1993) the formula of risk capital, per unit invested on net assets, can be approximated by the expression:

$$k^{MP} = 0.4 \cdot \sigma \sqrt{T} \tag{6.3}$$

where T denotes the time to maturity of the related option contract. This relationship is depicted in Fig. (6.1) as well, assuming the value $T = 1$.

Therefore, while the capital principle of Merton intersects the capitals axe at the origin and have a constant slope (equal to 0.4), the optimal principle of Eq. (6.1) intersects the capital axe at the value $k = -\mu$ and its slope depends on the liquidity premium ν.

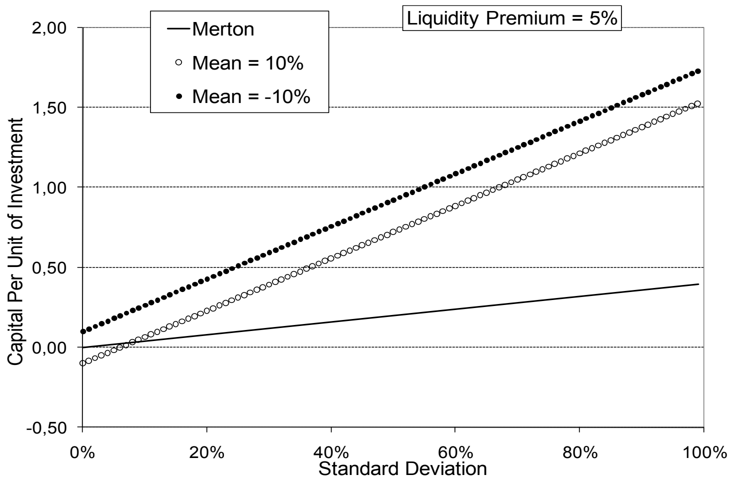


Figure 1: The Optimal Capital Line and the allocation rule proposed by Merton and Perold (1993).

This means, in particular, that at low volatilities those firms that obtain capital profits in average (characterised by $\mu > 0$) prefer to lend all their balances and do not maintain reserves at all. In fact, they only demand capital when:

$$\sigma \cdot \Phi^{-1}(1-\nu) - \mu > 0 \quad \Leftrightarrow \quad \sigma > \frac{\mu}{\Phi^{-1}(1-\nu)} \tag{6.4}$$

The point at which firms maintains a balance equal to zero, as predicted by the MM-proposition, is actually characterised by the condition:

$$\sigma = \frac{\mu}{\Phi^{-1}(1-\nu)} \tag{6.5}$$

By contrast, those firms that obtain capital losses in average (characterised by $\mu < 0$) always prefer to maintain some stock of reserves. According to the optimal capital principle, such stock is equal to the absolute value of the mean loss of net assets plus a term depending on the underlying volatility and the liquidity premium.

Therefore, the optimal capital principle determines an optimal capital line in the Gaussian case, which relates the optimal proportion of reserves in terms of the mean return and the volatility of the underlying risk, as well as the premium for liquidity offered in the market.

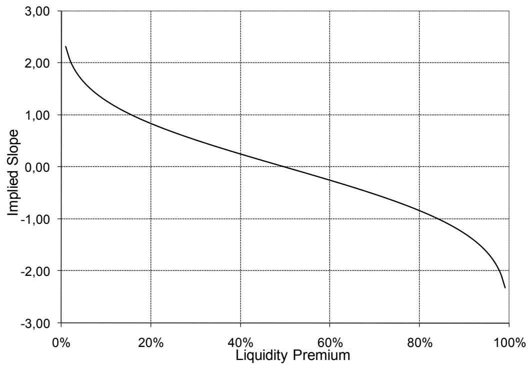


Figure 2: Implied Slope of the Optimal Capital Line in terms of the Liquidity Premium.

In fact, lower liquidity premiums are corresponded to markets where firms hold less cash in excess and are thereby willing to exchange more capital for every unit of increased volatility. Hence, as depicted in Figure 2, the slope of the optimal capital line is higher in such markets.

In particular, the slope tends to $+\infty$ when $\nu \to 0$, i.e. the capital line becomes vertical in this case. Hence the willingness to exchange balances for volatility can be regarded as unbounded in such markets ─ or in other words, the willingness to maintain cash balances cannot be satisfied in such markets.

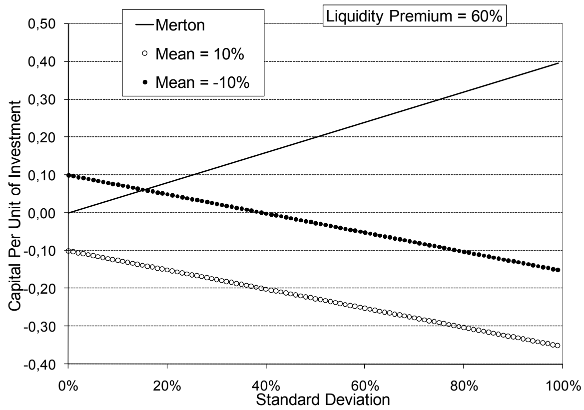


Figure 3: The Optimal Capital Line with negative slope.

Notice, however, still from Figure 2, that the slope of the capital line turns negative when $\nu > 0.5$. Accordingly, and contrary to the common intuition, the capital requirements may decrease with the level of volatility if the liquidity premium is sufficiently high.

This implies that when $\nu > 0.5$ firms eventually prefer to lend all their balances, once the underlying volatility surpasses a certain level determined by the expected return and the liquidity premium. Consequently, as depicted in Figure 3, the capital line takes negative values after some level of volatility is surpassed.

Finally, the Merton's principle can be obtained as a particular case of the optimal capital principle. Indeed, the rule of Eq. (6.3) can be obtained from Eq. (6.1) by replacing in Eq. (6.1) the values $\mu = 0$ and $\nu \approx 34.25\%$ (in such a way that $\Phi^{-1}(1-\nu) \approx 0.4$). See Figure 4.

This situation can be naturally corresponded to a competitive environment, where firms obtain no capital profits in average (as long as $\mu = 0$) and the liquidity premium is sufficiently high (around the level 34.25%).

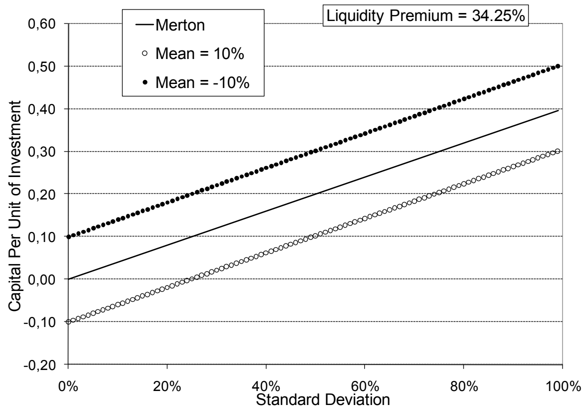


Figure 4: The Capital Rule of Robert Merton as a particular case of the Optimal Capital Principle.

Recall that the optimal capital principle, as defined according to Eqs. (5.7), (5.8) or (6.1), has been already obtained as an extension of the option-based Merton's principle. Thus, both principles relate the demand for capital to a demand for deposit insurance, in such a way that in both cases the maturity payments of the underlying net portfolio are described by Eq. (3.5). The difference relies on the fact that while in the model of Merton the underlying contingent claims are measured by the market prices of their corresponding call and put options, in the model of optimal capital proposed in this paper the same claims are assigned their actuarial prices.

We can then conclude that the Gaussian capital principle satisfactorily extends the model of capital proposed by Robert Merton ─ or equivalently, the principle satisfactorily extends the MM-proposition. The optimal capital principle is thereby meaningful from the economic point of view.

7. The Optimal Capital Principle with Heavy Tails

An appealing feature of the capital principle defined in Eqs. (5.7) and (5.8) is its adaptability to any family of probability distributions.

As a consequence, the capital requirements of different types of risks can be described on the same basis and hence, the model can be also implemented in institutions that hold securities exposed to non-homogeneous risks. This is particularly the case of insurance companies that simultaneously deal with highly standardised policies, such as car and fire insurance, as well as some individual contracts involving payments depending on events of low probability.

This is also the case of some financial conglomerates that simultaneously hold standard financial securities, transacted in highly liquid markets, as well as non-liquid derivatives and claims contingent on disaster and catastrophic events.

Two probability distributions that have raised a lot of attention in the Risk Management literature are the Exponential and the Paretian distributions, because they provide a mathematical description of the phenomenon of heavy tails.

Indeed, let the functions T^{EXP} and T^{PAR}, defined below, respectively denote the tail probability functions of Exponential and Paretian random variables:

$$T^{EXP}(x) = \exp\left(\frac{C-x}{\beta}\right) \quad \forall x \geq 0 \tag{7.1}$$

$$T^{PAR}(x) = \left(\frac{x}{B}\right)^{1/\alpha} \quad \forall 0 < x < B \tag{7.2}$$

Recall that the tail probability function of Gaussian random variables is expressed as:

$$T^{GAU}(X) = 1 - \Phi\left(\frac{\mu + x}{\sigma}\right) \quad \forall x \qquad (7.3)$$

Then the probability accumulated in the tails of Paretian risks is always greater than the probability accumulated in the tails of Exponential risks, which in turn is always greater than the probability accumulated in the tails of Gaussian risks:

$$T^{PAR}(x) > T^{EXP}(x) > T^{GAU}(x) \quad \forall x \qquad (7.4)$$

It is accordingly stated that Paretian tails are uniformly greater or heavier than Exponential tails, which in turn are uniformly greater or heavier than Gaussian tails.

In the actuarial literature, the condition (7.4) determines an order over a class of random variables that is known as the first stochastic order; see Goovaerts *et al.*, 1984. In other words, condition (7.4) implies that Gaussian risks are stochastically dominated by Exponential risks, and Exponential risks are stochastically dominated by Paretian risks.

From Eqs. (5.7), (7.1) and (7.2), the optimal capital principle under Paretian and Exponential risks is respectively given by:

$$k^{EXP}(\nu) = C - \beta \cdot \ln(\nu) \qquad (7.5)$$

$$k^{PAR}(\nu) = B \cdot \nu^{\alpha} \qquad (7.6)$$

The Gaussian capital principle $k^{GAU}(\nu)$, on the other hand, has been defined in Eq. (6.1).

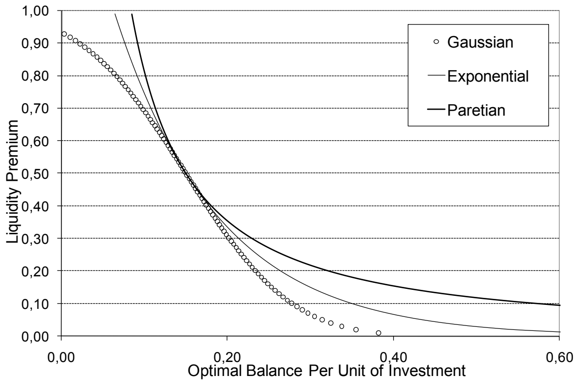


Figure 5: The Optimal Capital Principle for different risk parameterisations.

Consequently, as depicted in Figure 5, given any level of the liquidity premium, the Paretian capital principle is always greater than the Exponential capital principle, which in turn is always greater than the Gaussian capital principle.

Therefore, the optimal capital principle consistently assigns higher surpluses to riskier claims and hence, it is strictly risk-based.

8. Concluding Remarks

Firms that continuously trade capital and securities demand no cash reserves, for they can fit their balances at any moment through borrowing and lending [Modigliani and Miller, (1958)].

In fact, as proved by Merton (1974, 1977, 1997), although the market prices of equity and deposit insurance (which are the main components of the capital structure) actually depend on the level of reserves, the value of firms does not depend on it [see also Miller, (1998)]. This result is a consequence of the fact that continuous hedging suppresses risk.

However, when firms face restrictions on liquidity (in other words, when borrowing and lending may change the price of capital if the transacted amounts break on through certain thresholds) the capital structure determines the agency costs and the moral-hazard implicit in the contracts that managers respectively establish with stockholders and customers [Jensen, (1986), Froot *et al.*, (1993)].

Hence the market prices of equity and deposit insurance should be determined on an actuarial basis. As demonstrated in this paper, an optimal cash balance then exists, which leads to an optimal capital principle that is consistent with economic fundamentals and is easy to implement for a wide class of probability distributions.

Moreover, since the level of capital is explicitly related to the deductible or retention of the corresponding insurance contract, it explicitly represents the moral-hazard on the side of customers.

In particular, when the underlying risk follows a Gaussian probability distribution, a capital line is obtained relating the optimal proportion of capital to the standard deviation. This principle naturally extends the capital allocation rule proposed by Merton and Perold (1993).

9. References

[1] Basel Committee on Banking Supervision, (2004), *International Convergence of Capital Measurement and Capital Standards*, technical report. Available at http://www.bis.org

[2] Williams, J.R., Haka, S.F., Bettner, M.S. and Meigs, R.F., (2002), *Financial and Managerial Accounting: The Basis for Business Decisions*, McGraw-Hill.

[3] Merton, R.C., (1974), *On the Pricing of Corporate Debt: The Risk Structure of Interest Rates*, in: *The Journal of Finance*, 29 (2), The American Finance Association, pp. 449 – 470.

[4] Merton, R.C., (1977), *An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees*, in: *Journal of Banking and Finance*, 1, Elsevier, pp. 3 – 11.

[5] Modigliani and Miller, (1958), *The Cost of Capital, Corporation Finance and the Theory of Investment*, in: *The American Economic Review*, 48 (3), American Economic Association, pp. 261 – 297.

[6] Merton, R.C., (1997), *A Model of Contract Guarantees for Credit-Sensitive*, *Opaque Financial Intermediaries*, in: *European Finance Review*, 1 (1), Springer Netherlands, pp. 1 – 13.

[7] Ross, S.A., (1989), *Institutional Markets, Financial Marketing, and Financial Innovation*, in: *The Journal of Finance*, 44 (3), The American Finance Association, pp. 541 – 556.

[8] Cummins, J.D., Sommer, D.W., (1996), *Capital and Risk Property-Liability Insurance Markets*, in: *Journal of Banking and Finance*, 20, Elsevier, pp. 1069 – 1092.

[9] Miller, M.H., (1998), *The Modigliani-Miller Proposition After Thirty Years*, in: *Journal of Economic Perspectives*, 2 (4), American Economic Association, pp. 99 – 120.

[10] Goovaerts, M.J., Van den Borre, E., and Laeven, R.J.A., (2005), *Managing Economic and Virtual Economic Capital within Financial Conglomerates*, in: *North American Actuarial Journal*, 9 (3), Society of Actuaries, pp. 77 – 89.

[11] Venter, G., (1993), *Premium Calculation Implications of Reinsurance without Arbitrage*, in: *Astin Bulletin*, 21 (2), Peeters, pp. 223 – 230.

[12] Hull, J.C., (2000), *Options, Futures and other derivatives*, Prentice Hall.

[13] Goovaerts, M.J., De Vylder, F., Haezendonk, (1984), *Insurance Premiums*, Amsterdam: North-Holland.

[14] De Finetti, B., (1975), *Theory of Probability: A Critical Introductory Treatment*, Willey.

[15] Froot, K.A., Scharfstein, D.S., and Stein, J.C., (1993), *Risk Management: Coordinating Corporate Investment and Financing Policies*, in: *The Journal of Finance*, 48 (5), The American Finance Association, pp. 1629 – 1658.

[16] Jensen, M.C., (1986), *Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers*, in: *The American Economic Review*, 76 (2), American Economic Association, pp. 323 – 329.

[17] Fama, E., (1980), *Agency Problems and the Theory of the Firm*, in: *Journal of Political Economy*, 88 (2), The University of Chicago Press, pp. 288 – 307.

[18] Dhaene, J., Goovaerts, M.J., and Kaas, R., (2003), *Economic Capital Allocation Derived from Risk Measures*, in: *North American Actuarial Journal* 7 (2), Society of Actuaries, pp. 44 – 59.

[19] Laeven, R.J.A., Goovaerts, M.J., (2004), *An Optimization Approach to the Dynamic Allocation of Economic Capital*, in: *Insurance: Mathematics and Economics*, 35 (2), Elsevier, pp. 299 – 319.

[20] Mierzejewski, F., (2006), *Optimal Capital Allocation Confronting Bankruptcy and Agency Costs*, in: *Bank- en Financiewezen*, 2006/2, Larcier, pp. 72 – 77.

[21] Merton, R.C., Perold, A.F., (1993), *Theory of Risk Capital in Financial Firms*, in: *Journal of Applied Corporate Finance*, 6 (3), Morgan Stanley, pp. 16 – 32.

LA RESPONSABILITÉ SOCIALE ET L'ÉTHIQUE ENVIRONNEMENTALE – ÉLÉMENTS DE LA CULTURE ÉCOLOGIQUE DANS LES ENTREPRISES DU XXIe SIÈCLE: L'ETUDE SUR LES ENTREPRISES ROUMAINES

Marilena **MIRONIUC**
Faculté d'Economie et d'Administration des Affaires
Université "Alexandru Ioan Cuza", Iaşi, **ROUMANIE**

Abstract:

During the last decade, the business community ends especially the companies that activate in the fields responsible for the degradation of the environment, are seen as subjects whose main role is to ensure sustainable development. They must face growing pressures from the stakeholders, who wish to know their position regarding sustainable development, in its three dimensions: economical development, that generates profits and jobs; social development, as a guarantee for the well-being and for the compliance to human rights; environmental development, which must ensure the preservation of natural resources and the ability of the ecosystem to absorb and tolerate pollution. A company is defined as sustainable if it is socially responsible. This paper is the result of a process of investigation, of an analysis of the present framework, and of a bibliographical synthesis in the field of social responsibility and environmental ethics. First of all, we have made a historical incursion in the problematic of social responsibility of companies, and then we have mentioned the most important challenges that justify socially responsible behavior in today's world. In order to point out the degree of responsibility of Romanian enterprises, regarding the actions for the protection of the environment, we have drawn an empirical analysis on a representative sample of Romanian companies.

Keywords: corporate social responsibility, pro-social responsibility culture

JEL classification: M14, Q5

1. Incursion historique dans la problématique de la responsabilité sociale de l'entreprise: une revue de la littérature

Les racines de la **responsabilité sociale de l'entreprise** (RSE) ne sont récentes, même si ce concept a commencé à attirer l'attention surtout à partir des premières décennies du 20[e] siècle. Le comportement responsable est né au moment où les gens ont commencé à développer des relations du type économique, mais pendant les dernières années cette pratique a été reconnue comme nécessaire et compatible avec le but de l'entreprise d'obtenir du profit et du bien-être social.

Pendant les années '50, l'expression *Corporate Social Responsibility* était présente dans la doctrine des entreprises des Etats-Unis, étant liée au thème social et à la relation de l'entreprise avec l'environnement [Marchetti, (2006)]. En cette période, les grandes compagnies américaines étaient accusées de certaines pratiques antisociales et on a essayé d'amoindrir leur pouvoir par des lois antitrust, par des actes normatifs bancaires et par des lois qui protégeaient les consommateurs [Bianchini, (2004)]. Dans ces conditions, les hommes d'affaires étaient obligés de respecter ces polices et d'adopter seulement les décisions acceptées par la société, en termes de valeurs et d'objectifs. Ainsi est survenue la nécessité que les entreprises assument des obligations sociales en parallèle avec leurs fonctions économiques, de production et de commercialisation, orientées vers l'obtention du profit. Déjà, à partir de cette période, les entreprises étaient vues comme des sujets qui jouaient un rôle important pour garantir le progrès social et cultural, aussi bien que pour améliorer le bien-être de la société civile.

Dans les années '70, les discussions sur la « *réceptivité sociale de l'entreprise* » se sont intensifiées, surtout concernant la manière dont les entreprises peuvent s'adapter aux besoins de la société et aussi concernant la « *performance sociale de l'entreprise* », c'est-à-dire les résultats des actions de responsabilisation sociale initiées par les entreprises [Stancu and Orzan, (2006)]. Dans la même période est apparue une série d'idéologies de l'environnement, à mesure que les sociétés industrialisées sont devenues conscientes de la gravité des problèmes liés à l'environnement. Ainsi, pour la première fois, en 1972, lors de la Conférence sur l'Environnement de Stockholm, ont été mises les bases du développement d'une politique environnementale au niveau international [Turner, Pearce, and Bateman, (2003)].

Pendant la décennie suivante, la théorie des *stakeholders* (*stakeholders theory*) apporte des contributions importantes concernant la responsabilité sociale de l'entreprise. R.E. Freeman, le représentant principal de cette théorie, change le paradigme des *stakeholders* de Milton Friedman, qui considérait que « la seule responsabilité sociale des affaires est d'utiliser leurs ressources et de s'impliquer dans des activités visées à augmenter les profits, tout en respectant les règles du jeu, pratiquant une concurrence ouverte et libre, sans tromperies ni fraudes » [Friedman, (1962)]. L'augmentation de la valeur pour les actionnaires et sa maximisation représentaient, dans l'opinion de cet auteur, la plus grande responsabilité de l'entreprise. Mais la théorie fondée par R.E. Freeman affirme que les entreprises ont un but public plus large: celui de créer une valeur pour la société, comportant des responsabilités multiples envers tous les *stakeholders* qui interagissent avec l'entreprise [Freeman, and Werhane, (1999)]. Les intérêts parfois opposés des *stakeholders* doivent être réconciliés, dans l'opinion de l'auteur, si l'on veut que l'entreprise survive et se développe**.** Le rôle majeur dans la réconciliation des parties revient au dirigeant qui, grâce à son pouvoir décisionnel et à l'accès aux informations, cherche à maximiser une fonction d'utilité multicritère, remplie par lui-même aussi bien que par les sujets impliqués dans l'activité de l'entreprise, sous la restriction de l'existence d'un profit minimum, absolument nécessaire pour préserver la position du dirigeant. Il était nécessaire de repenser les modèles de gestion des grandes entreprises et de passer des modèles de responsabilité monocratique à des modèles de responsabilité des parties prenantes multiples. Dans la même période, l'approche des parties prenantes et le concept de *Business Etichs* – l'éthique des affaires, conformément auquel on attribuait une signification morale aux actions individuelles et collectives des partenaires d'affaires et aux conséquences sociales de l'activité des entreprises, s'imposaient comme des éléments constants de la gestion stratégique des entreprises. En plus, la responsabilité des opérations économiques qui se reflètent sur le plan social était vue par les entreprises comme représentant une opportunité pour les affaires (Marchetti S., 2006).

La théorie des *stakeholders* montre ses limites à la fin des années '80 et au début des années '90, quand on a constaté qu'il est difficile de construire une affaire prospère, responsable, qui ne génère pas un impact négatif sur les individus et sur l'environnement. Dans ce contexte, la Commission Mondiale pour l'environnement et le développement (*World Commission on Environment and Development*), appelée aussi la Commission Brundtland des Nations Unies (1987), met les bases du concept du développement durable, qui signifie « *le développement qui satisfait les besoins de succès de la génération présente sans compromettre les attentes des générations futures de satisfaire leurs propres besoins* » (www.europa.eu.int, 2004). En essence, cette définition est basée sur le concept de « nécessité », qui nous fait voir le développement économique en tant que moyen d'amélioration de la qualité de la vie. En 1992, à Rio, au cadre de la Conférence des Nations Unies pour l'environnement et le développement, ont été élaborés les « principes de Rio » (Agenda local 21) qui indiquent *le développement durable au 21^{e} siècle*. Le forum de Rio a influencé largement le destin de l'humanité, en promouvant un nouveau concept de développement, qui inclut trois dimensions également importantes (la *Triple Bottom Line*): *le développement économique*, qui génère le profit et des emplois; *le développement social*, qui garantie le bien-être et le respect des droits de l'homme; *le développement environnemental*, qui doit assurer la protection des ressources naturelles et la possibilité de l'écosystème d'absorber et de tolérer la pollution. L'agenda 21 souligne la nécessité d'adopter à des niveaux différents (mondial, européen et national) des programmes d'action qui visent à atteindre un *développement durable dans une vision globale, par des actions et des processus de décision entrepris aux niveaux locaux*. Un rôle important est donc attribué aux autorités locales, qui devaient collaborer avec la communauté civile pour formuler des plans d'actions qui assurent le développement durable au niveau local. Pour la première fois, les entreprises sont vues comme des sujets au rôle principal pour assurer le développement durable. Le débats internationaux dans ce domaine ont conduit à l'établissement d'une interconnexion entre le développement durable et la responsabilité sociale de l'entreprise [Mauri, and Valentini, (2003)]. Une entreprise durable est considérée celle qui a une responsabilité sur le plan social. Cette responsabilité, avant d'être légale, a une nature morale envers tous ceux avec qui l'entreprise a des relations: à partir des clients/

consommateurs jusqu'aux générations futures, des fournisseurs aux salariés, des citoyens aux actionnaires, des créditeurs aux autorités de contrôle [Mironiuc, (2005)].

La Commission Européenne a eu l'initiative relevante d'élaborer une définition de la responsabilité sociale de l'entreprise, commune au niveau européen, publiée (2001) dans la Carte Verte de la Commission Européenne, intitulée « Promouvoir un cadre européen pour la responsabilité sociale des entreprises. » Ainsi, la responsabilité sociale de l'entreprise signifie l' « intégration volontaire des aspects sociaux et écologiques dans les opérations commerciales et dans les rapports de l'entreprise avec les partenaires des affaires intéressés ». [Stancu, and Orzan, (2006)]. Alors, la responsabilité sociale de l'entreprise représente un comportement volontaire du monde des affaires, non-imposé par la loi et en dehors des obligations législatives imposées par chaque pays, qui doit être intégrée dans la stratégie de l'entreprise pour garantir une manière responsable de gestion des affaires. En 2002, la Commission Européenne propose une stratégie de promotion de la gestion responsable, basée sur: la mise en évidence de l'impact positif de la responsabilité sociale sur l'environnement des affaires et sur la communauté en général, sur le développement de l'échange d'expériences et de bonnes pratiques concernant la responsabilité sociale et sur la promotion de la responsabilité sociale dans les petites et moyennes entreprises (PME). En mars 2006, la Commission Européenne s'est proposé de transformer l'Europe dans un pôle d'excellence de la responsabilité sociale des entreprises et a lancé « L'Alliance pour la RSE », en unifiant des entreprises responsables socialement déjà actives, dans le but de contribuer à la stratégie européenne de croissance économique et de création d'emplois.

Au-delà du discours européen commun, les pratiques de la responsabilité sociale de l'entreprise évoluent différemment d'un pays à un autre, en fonction du spécifique économique et social de chaque région et en fonction de la succession de l'adhérence des pays à l'Union Européenne. Des recherches récentes montrent que, à la différence de la doctrine américaine, qui a abordé la problématique de la responsabilité sociale en se concentrant sur les clients, considérés les principaux *stakeholders*, en Europe l'attention est accordée différemment aux problèmes éthiques et aux conflits moraux générés par les pratiques RSE (dans la Grande Bretagne), aux problèmes de l'environnement (dans les pays nordiques), aux aspects sociaux (dans les pays sud-européens) [Marchetti, (2006)].

2. Les raisons qui justifient le comportement socialement responsable dans le monde contemporain

La responsabilité sociale de l'entreprise représente une thématique qui s'est imposée dans l'attention de la société civile surtout comme une réaction à deux phénomènes : *la globalisation des systèmes économiques et la crise des marchés financiers* [Marchetti, (2006)], qui ont déterminé de nouvelles visions sur le développement économique et social.

2.1. L'Internationalisation de l'économie et la délocalisation des affaires

Après plus d'une décennie, on a constaté que la globalisation, malgré les attentes, ne représente pas une panacée. L'accès au *système des échanges internationaux* des pays moins développés semble être limité et le jeu de la concurrence globale risque d'augmenter dans ces cas le degré de pauvreté. Le marché global et l'extension du rôle des investissions du capital étranger dans la structuration de la nouvelle économie globale ne déterminent pas automatiquement une distribution plus équitable de la richesse, mais peuvent amplifier dramatiquement les inégalités économiques et sociales au niveau international [Bianchini, (2004)]. Les formes traditionnelles de redistribution, pratiquées par les gouvernements comme suite à la collecte des ressources internationales, sont déficitaires et insolvables dans les conditions de l'internationalisation de l'économie, parce que la richesse produite dans certains pays se déplace vers d'autres. Les compagnies multinationales qui ont le siège dans des pays développés mais qui produisent et achètent des matières premières dans des pays en cours de développement, utilisent les compagnies de ces pays en tant que fournisseurs indirects. Le niveau avantageux des coûts dans les pays en voie de développement, en comparaison avec celui des pays développés, est dû en partie à l'absence des standards sociaux et environnementaux requis dans les pays développés et aux normes de protection des parties prenantes (« *dumping social* ») [Sacconi, (2003)]. L'absence des réglementations dans ces domaines ou l'inefficacité de leur application représentent des conditions favorables pour les compagnies multinationales qui peuvent expatrier, d'une manière opportune, des parties signifiantes des bénéfices locaux générés par l'intermédiaire des transactions avec ces pays récemment intégrés dans le marché global. Il est certain qu'avec le transfert, pas toujours réussi, vers les pays en voie de développement de ce modèle de croissance économique il a eu lieu un transfert culturel aussi. L'implication des compagnies multinationales dans le développement des polices liées au bien-être, à la protection de l'environnement et aux droits des salariés dans les pays en cours de développement représente l'expression de leur comportement responsable, mais aussi un enjeu pour la réduction des pressions dans les pays d'origine des compagnies multinationales, par la réduction de la compétition par les coûts. La globalisation représente donc un processus accéléré d'intégration économique, dont les conséquences se reflètent dans le développement durable et demandent aux grandes entreprises d'introduire *des rectifications sociales* dans leur activité.

2.2. Les scandales financiers et le pouvoir discrétionnaire des dirigeants

Les années '90 ont marqué la période de la mondialisation, où les compagnies multinationales se sont largement développées. Après cette époque, le marché financier global a été marqué par des *crises et turbulences* graves, provoquées par la découverte de certains faits moins honnêtes des administrateurs de quelques compagnies cotées importantes, qui ont truqué des comptes et qui ont manipulé des informations réservées pour montrer des niveaux de performance qui soutiennent la valeur des titres, pour paraître intéressantes pour les investisseurs. Les crises mentionnées ont été en grande partie le résultat de l'abus de *pouvoir discrétionnaire des dirigeants* et des administrateurs, au détriment d'autres sujets. La technique des options de l'achat d'actions et la rémunération des dirigeants en fonction de leur performance n'ont pas pu parfaire l'alignement des intérêts des administrateurs délégués avec ceux des actionnaires, et ont provoqué de nombreux préjudices aux actionnaires minoritaires, qui n'appartenaient pas au groupe de contrôle des dirigeants, à tous les partenaires des affaires, aux consommateurs, à la communauté où fonctionnaient ces compagnies. La série noire débute par la faillite spectaculaire du gigantesque groupe économique Enron Corporation. Ce phénomène a aussi touché WorldCom (la faillite la plus signifiante de l'histoire des Etats-Unis), Tyco International, Qwest, Xerox, de la série américaine, Vivendi Universal, Ahold et Parmalat, en Europe [Feleagă, (2006)]. La série de ces scandales financiers a généré une crise de confiance dans la fiabilité des informations fournies par la comptabilité, qui ne peuvent pas surprendre toutes les dimensions de la performance et qui ont mis en évidence la dissociation entre la comptabilité, l'éthique et la responsabilité des dirigeants [Pillips, (2003)]. Aujourd'hui, le développement des marchés financiers et les demandes de communication financière requièrent *mesurer les performances dans une vision globale*, financière et non financière, y inclus par *des indicateurs qui reflètent la performance des pratiques de responsabilité sociale des entreprises*. En plus, les défis internationaux et les limites des règles législatifs de dépister et sanctionner les comportements incorrects ont sensibilisé beaucoup plus la conscience des parties prenantes concernant l'importance de la responsabilité sociale. La finalité des initiatives visant à consolider la responsabilité sociale de l'entreprise est donc de promouvoir « l'éthique des affaires. » Dans ces conditions, dans l'entreprise qui adopte la responsabilité sociale il se crée une réputation d'équité entre ceux qui opèrent à son intérieur, et ce comportement tend à devenir un trait distinctif de leur propre identité et un *élément de la culture de l'entreprise*.

2.3. Le changement du profil des consommateurs

L'évolution des technologies de l'information, respectivement la possibilité d'échanger des informations en temps réel à un coût bas, a permis à un nombre de plus en plus grand de personnes l'accès aux informations, et a demandé plus de transparence en ce qui concerne le comportement des entreprises. Cela a contribué au changement du profil des consommateurs potentiels. Le type de consommateur récepteur passif (*le consommateur client*) des offres de l'entreprise cède sa place à un sujet actif qui veut consommer d'une manière critique (*le consommateur citoyen*), qui veut connaître les coûts sociaux des produits /services qu'il consomme. L'apparition d'un segment de consommateurs sensibles à la thématique environnementale stimule le développement des produits et technologies innovateurs, qui ont un impact réduit sur l'environnement. La qualité écologique des produits est devenue une arme du succès commercial, par laquelle la concurrence au niveau international se sert des prestations de la politique environnementale pour se créer des avantages compétitifs, après une comparaison des avantages socio-culturaux et écologiques avec les points faibles remarqués comme suite des coûts supplémentaires induits par les industries propres et par les technologies anti-polluantes. Dans le même contexte il apparaît le concept de marketing écologique comme un instrument pour: la promotion des politiques commerciales d'importation, orientées vers le transfert des technologies écologiques, considérées comme des *inputs* qui comportent les compétences nécessaires pour la réduction ou la limitation du niveau de pollution de certaines zones; l'élargissement du marché des produits écologiques par la promotion des brevets et des licences et par les étiquettes écologiques; l'établissement des restrictions de l'exportation/ importation des déchets dangereux; l'attraction du capital étranger pour consolider la production nationale des biens et services écologiques [Cămăşoiu, Popescu, Cernat, and Pană, (2003)].

2.4. L'Optimisation inter générations

Le système économique contemporain, qui fournit tout le nécessaire d'un niveau moderne de vie, est un système où toutes les formes de capital ont une valeur, y inclus le capital naturel, sans l'aide duquel il ne pourrait pas fonctionner. Or, dans le système économique global, on constate une évolution inverse entre *le capital naturel qui diminue*, et *le capital humain qui abonde*. Au niveau international, les opinions concernant les causes de la dégradation de l'environnement sont nombreuses et en dynamique, mais la plupart convergent vers la considération que la pollution provient des pressions déterminées par la *croissance de la population* et par le niveau *de la consommation des ressources* relativement limitées de l'écosystème. L'interaction entre la croissance de la population et le développement économique, l'indisponibilité des ressources naturelles et l'incapacité de l'environnement à absorber les effets nuisants, irréversibles et cumulatifs de la production et de la consommation sont considérées par beaucoup d'environnementalistes les raisons de la « croissance zéro ». L'humanité a été jusqu'à un moment donné ignorante, puis incapable de mobiliser de l'intelligence et des ressources pour rétablir l'équilibre écologique. Dans la vision du développement durable, il faut une optimisation *inter et intra générations*, par la promotion d'un esprit de solidarité, par des soins réciproques, par la tolérance pour ne pas diminuer les possibilités des générations présentes et futures d'utiliser les ressources naturelles nécessaires pour maintenir la qualité de la vie. Par conséquent, les générations présentes doivent transmettre aux générations futures un stock de ressources qui ne soit pas inférieur à celui qu'elles-mêmes possèdent (« *la règle du capital constant* »). Mais, sachant que les ressources naturelles représentent un capital qui, s'il diminue, il ne peut pas être parfaitement remplacé par d'autres formes de capital physique, parce qu'il remplit des fonctions essentielles pour la survie de l'espèce humaine, elles doivent être préservées (« *la règle du capital naturel critique* »).

Le placement des intérêts de la génération présente devant ceux des générations futures ne doit pas sacrifier les premiers, mais plutôt affirmer des principes éthiques, sachant que le marché libre fonctionne d'après le principe de l'allocation des ressources au plus performant, qui sait comment maximiser leur rendement. Conformément à la littérature de spécialité occidentale [Turner, Pearce, and Bateman, (2003)], l'assurance de l'éthique envers les générations futures suscite la réponse à la question suivante: « *Est-il juste que ceux qui sont en vie à présent exploitent, par des raisons économiques, les ressources environnementales jusqu'à les détruire, en en tirant des avantages et en transférant des coûts vers les générations à venir, qui ne sont pas encore nées et qui ne peuvent pas exprimer leur propre opinion à ce propos ?* » Etant donnée la nature humaine égocentriste, les chercheurs du développement durable essaient de déduire combien du comportement humain doit changer pour que l'objectif de l'équité inter et intra générations soit atteint. Certains chercheurs [Poli, (1994)] affirment que « *l'éthique de la bonne administration* » suffit pour garantir le développement durable. De toute manière, l'éthique environnementale implique un contrat social entre les générations, par lequel les générations présentes assument des obligations vis-à-vis des générations futures. L'éthique environnementale est caractérisée par le sens de la responsabilité et est liée au respect des droits de l'homme. Dans les conditions où *la responsabilité environnementale* a un rôle fondamental pour l'adoption des décisions de nature politique, économique et sociale, le développement à long terme d'une *culture écologique* devient une nécessité dans le monde moderne.

2.5. L'Economie du savoir et l'innovation

Pendant les deux dernières décennies, l'économie des pays développés s'est dématérialisée. Elle est passée d'un système économique où les avantages compétitifs dépendaient surtout des ressources matérielles et financières gérées et contrôlées par les entreprises, à un système économique (*knowledge economy – l'économie du savoir*) où les performances sont conditionnées, de plus en plus, par le complexe des ressources immatérielles créées par l'activité antérieure des entreprises ou par celle à laquelle elles ont accès. Le système économique global est en cours de devenir un système des « idées et innovations technologiques » où le potentiel du capital intellectuel devient un levier compétitif pour la survie économique [Bianchini, (2004)]. Chaque jour fait croître d'une manière signifiante l'influence du « capital intellectuel » de l'entreprise, représenté par : *des connaissances et des habiletés d'ordre organisationnel* (le capital organisationnel)*, la loyauté et la cohésion du personnel* (le capital humain), *la crédibilité de l'entreprise* (le capital relationnel). Le marché boursier accorde à ces ressources, dont la formation requiert des efforts considérables, une valeur de plus en plus élevée. Le prix des actions augmente quand les ressources intangibles d'une entreprise ne diminuent pas en ce qui concerne leur valeur et ne sont pas reproduites par les entreprises rivales. Dans le contexte de la demande globalisée de main d'oeuvre, les salariés sont de plus en plus conscients qu'ils représentent « le bien le plus valeureux » de l'entreprise, ayant des compétences et des habiletés qui appartiennent en premier lieu à chaque individu, et seulement après à l'entreprise. Les contrats de travail donnent le droit à l'entreprise de contrôler d'une manière limitée ces compétences, et la mobilité du personnel sollicite la prudence des dirigeants lorsqu'ils veulent baser leur stratégie sur des compétences spécifiques du personnel. Les études montrent que la valeur générée par le capital humain des entreprises dépasse souvent 50% de la valeur créée par le capital intellectuel, tout comme la perte partielle du capital humain peut avoir des conséquences catastrophiques sur les avantages compétitifs et sur la valeur sur le marché de l'entreprise [Del Bello, and Gasperini, (2006)]. Il est évident que les changements qui ont lieu dans la société globale demandent aux entreprises d'être responsables du point de vue social et préoccupées par l'excellence, par l'augmentation de la valeur sociale et par des investissements en l'intelligence, dans le contexte plus large de la gestion du savoir. En plus, les entreprises tendent à devenir les principaux protagonistes dans la relation avec la société: en investissant dans leur réputation sur le marché, par la confiance des consommateurs, par la qualité des relations avec les salariés.

Puisque sur les marchés concurrentiels de plus en plus internationalisés on a constaté, pendant les dernières années, une dissociation entre l'éthique et la responsabilité des affaires, l'équilibre entre les trois dimensions (économique, environnementale et sociale) est difficile à accomplir sans entrer dans le jeu de la responsabilité sociale de l'entreprise, qui n'est pas un phénomène passager, lié à une mode culturelle, mais un élément permanent, destiné à définir le comportement des entreprises au 21^{e} siècle [Bianchini, (2004)].

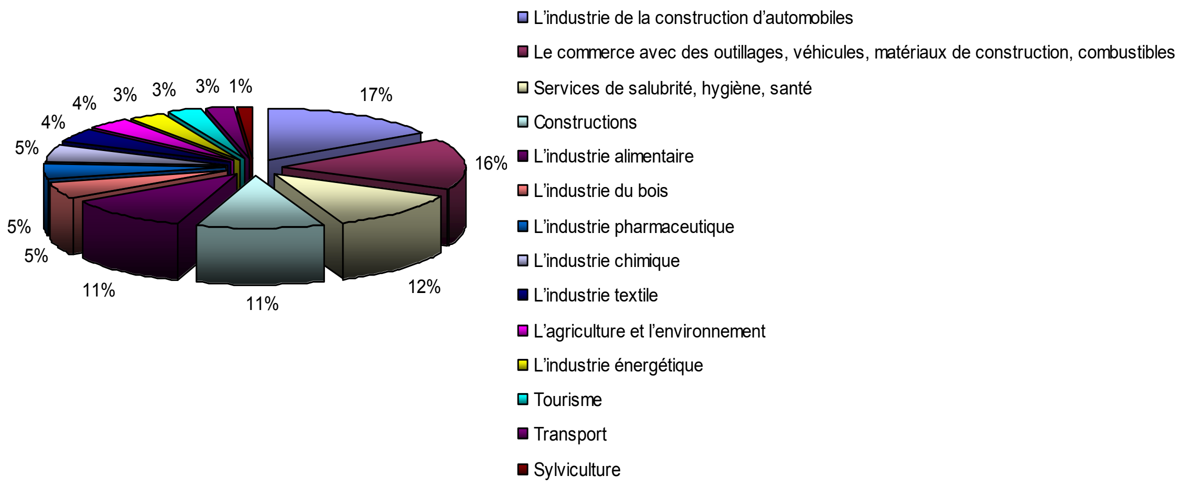
3. Le début de la responsabilité sociale de l'entreprise en Roumanie – Etude empirique

La responsabilisation sociale des entreprises vient d'apparaître en Roumanie. La sensibilité vers les variables sociales et environnementales, tout comme le contrôle permanent de la relation entre l'économie et l'environnement représentent de nouveaux repères dans l'évaluation des performances dans l'époque contemporaine. Il existe des études qui montrent que reconnaître qu'une entreprise est responsable socialement a des effets de rentabilité et des effets relationnels, qui peuvent créer des avantages compétitifs certains. Les entreprises roumaines doivent adapter leur stratégie d'affaires en rapport avec ces changements, réévaluer la durabilité des actions entreprises, en mesurant et en gérant leur propre impact social, environnemental et économique.

3.1. Les hypothèses et la méthodologie de la recherche

Pour apprécier le degré de responsabilisation sociale des entreprises de notre pays et leur perception, surtout quant à la dimension environnementale du développement durable, nous avons fait une analyse empirique sur un échantillon de 157 entreprises, pour la plupart de la région du Nord-Est de la Roumanie. La structure de l'échantillon est formée de 6,37% de sociétés cotées à la Bourse des valeurs de Bucarest dans la catégorie I et 93,63% d'entreprises non cotées (PME), dont 51,70% micro entreprises et 48,30% moyennes entreprises. Même si les principes de la responsabilité sociale ont été fondés pour être appliqués dans les compagnies multinationales qui opèrent dans un système capitaliste avancé, le grand poids des micro entreprises qui entrent dans l'échantillon analysé ne peut pas représenter un point faible, étant donnée la préoccupation au niveau international d'adapter les principes et les pratiques de la responsabilité sociale aux PME. Du total des sociétés cotées dans la catégorie I ont été exclues les sociétés de services d'investissements financiers et les banques, parce que par leur activité elles n'ont aucune liaison directe avec les problèmes de l'environnement, mais d'une manière indirecte, par le financement des investissements éthiques pour la protection de l'environnement. Pour tirer des conclusions aussi réalistes que possible, l'échantillon a été formé d'une mosaïque d'entreprises qui appartiennent à des secteurs d'activité qui ont un impact différent sur l'environnement (Figure 1).

Figure 1: La structure de l'echantillon



Les données nécessaires à l'analyse sont relatives à la période 2006 – 2007 et ont été cueillies dans le cas des sociétés cotées par l'accès direct aux rapports financiers annuels publiés sur l'Internet, et dans le cas des autres sujets, par l'administration d'un questionnaire, directement ou par e-mail. Le questionnaire était adressé au chef d'entreprise ou à un cadre supérieur.

Le questionnaire, composé de neuf questions, a été conçu de manière à évaluer, pour la période analysée: l'intérêt des répondants pour les investissements concernant l'environnement (la prévention et le lutte contre la pollution, la protection des ressources naturelles et la préservation de la biodiversité, la création/ promotion des produits écologiques); le poids des investissements dans l'environnement dans la chiffre d'affaires; la perception des investissements dans l'environnement comme une contrainte qui détermine des coûts ajoutés ou par contre comme une opportunité; les dépenses liées à l'environnement, qui affectent l'activité de l'entreprise (autorisations/ permis d'environnement, dépenses pour la formation du personnel dans le domaine de l'environnement; amendes pour des non conformités liées à l'environnement; le couvrement des préjudices causés par la pollution); l'existence dans les entreprises du personnel spécialisé dans le domaine de l'environnement; la disponibilité des directeurs de créer des emplois dans le domaine de la gestion et de la protection de l'environnement; la hiérarchisation de dix critères de performance en fonction de leur importance pour les perspectives des entreprises et pour la communication financière. Le questionnaire était adressé au chef d'entreprise ou à un cadre supérieur. Le traitement des données a utilisé le programme SPSS (*Statistical Package for the Social Sciences* – Le Paquet statistique pour les sciences sociales).

L'étude empirique s'est proposée de vérifier quelques hypothèses reconnues dans la littérature de spécialité (Tableau 1):

Tableau 1. Tableu croisé des questions et des hypothèses testées dans le questionnaire

	Questions	1	2	3	4	5	6	7	8	9
Hypothèses	1. Les petites et moyennes entreprises adoptent, avec quelques exceptions, un comportement passif en ce qui concerne les initiatives sociales et environnementales		X	X		X	X	X		
	2. Les petites et moyennes entreprises perçoivent l'environnement comme une contrainte qui cause des coûts ajoutés		X	X	X					
	3. Les initiatve sociales représentent un des outils le moins souvent communiqué et utilisé par les dirigeants pour suivre et anticiper la performance, par raport aux indicateurs comptables et financiers traditionnels								X	

3.2. Resultats

En synthèse, nous mentionnons quelques résultats de cette étude. Dans le cas de chaque hypothèse, nous décrivons les questions-test et leurs résultats respectifs. Les questions 1 et 9 insérées dans le questionnaire ont permis la structuration de l'échantillon en fonction de la taille des entreprises (le nombre des salariés, le chiffre d'affaires anuală netă, l'actif total) et du secteur d'activité auquel elles appartiennent. Les questions qui testent plusieurs hypothèses sont décrites une seule fois.

Hypothèse 1: *Les petites et moyennes entreprises adoptent, avec quelques exceptions, un comportement passif en ce qui concerne les initiatives sociales et environnementales* est testé, directement ou indirectement, par les questions 2, 3, 5, 6 et 7. Nous présentons seulement les résultats des questions 5, 6 et 7 qui testent l'hypothèse directement.

La question 5 - Pendant les deux dernières années l'activité de l'entité a été grevée par des dépenses pour: **a.** des autorisations/ permis environnementaux; **b.** la formation du personnel dans le domaine de la gestion de l'environnement; **c.** amendes pour le manque de respect des normes environnementales (précisez le montant); **d.** couvrir les dommages causés par la pollution.

Les dépenses liées à l'environnement qui grèvent l'activité des entreprises analysées sont représentées dans le cas de 47,14% des sujets questionnés concernant les dépenses pour l'obtention des autorisations environnementales (39 micro entreprises et 36 moyennes entreprises), 13,40% des répondants font des dépenses pour la formation du personnel dans le domaine de la gestion de l'environnement (10 sociétés cotées, 11 moyennes entreprises), 2,56% ont couvert des dommages causés par la pollution (1 société cotée, 3 moyennes entreprises), 0,63% des sujets ont supporté des amendes pour ne pas avoir respecté les normes environnementales (1 moyenne entreprise), la différence jusqu'à 100% étant représentée par des dépenses combinées, pour deux ou trois des cas mentionnés (Tableau 2).

Tableu 2. La structure des dépenses pour des actions de l'environnement

No.	Dépenses pour:	Entreprises %	No.
1.	Des autorisations/ permis environnementaux	47,14	74
2.	La formation du personnel	13,40	21
3.	Amendes	0,00	0
4.	Couvrir les dommages causés par la pollution	2,56	4
5.	Des autorisations/ permis et la formation du personnel	12,10	19
6.	Des autorisations/ permis et couvrir les dommages causés par la pollution	1,91	3
7.	Des autorisations/ permis et amendes	0,63	1
8.	Des autorisations/ permis, la formation du personnel et amendes	0,63	1
9.	La formation du personnel etcouvrir les dommages	0,63	1
10.	Sans réponse	21,00	33
	Total	**100**	**157**

La question 6 - Disposez-vous de personnel spécialisé avec des attributions dans le domaine de la gestion et de la protection de l'environnement ? **a.** oui (précisez combien); **b.** non (Tableau 3).

Tableau 3. La structure des entreprises selon le personnel spécialisé en gestion de l'environnement

No.	Personnel spécialisé en gestion de l'environnement:	Entreprises %	No.
1.	Oui	40,76	64
2.	Non	59,24	93
	Total	**100**	**157**

La question 7 - Avez-vous créé pendant les deux dernières années des emplois dans le domaine de la gestion et de la protection de l'environnement ? **a.** oui (précisez combien); **b.** non.

Le personnel spécialisé, qui a des attributions dans le domaine de l'environnement, existe dans le cas de 40,76% du total des entreprises questionnées, respectivement 64 entreprises (10 sociétés cotées et 54 moyennes entreprises) disposent d'un nombre de 71 personnes ayant des compétences dans le domaine de la gestion et de la protection de l'environnement, dont 53 postes ont été créés entre 2006 – 2007, surtout dans les entreprises cotées en Bourse. De même, 24 de moyennes entreprises n'ont qu'un seul employé spécialisé dans ce domaine. Un coefficient de pondération de 59,24% du total des sujets questionnés est détenu par un nombre de 93 micro entreprises qui n'ont pas de personnel spécialisé dans le domaine de la gestion et de la protection de l'environnement. Les résultats obtenus confirment l'hypothèse no. 1.

Hypothèse 2: *Les petites et moyennes entreprises perçoivent l'environnement comme une contrainte qui cause des coûts ajoutés* est testée par les questions 2, 3 et 4.

La question 2 - Dans les deux dernières années, vous avez fait des investissements pour: **a.** prévenir et combattre la pollution; **b.** la protection des ressources naturelles et la préservation de la biodiversité; **c.** la création et la promotion des produits/ services écologiques.

La question 3 - A quel pourcentage du chiffre d'affaires ont remonté ces investissements ? **a.** moins de 5%; **b.** entre 5-10%; c) plus de 10% (Figure 3).

Figure 2: La structure des investissements concertant l'environnement

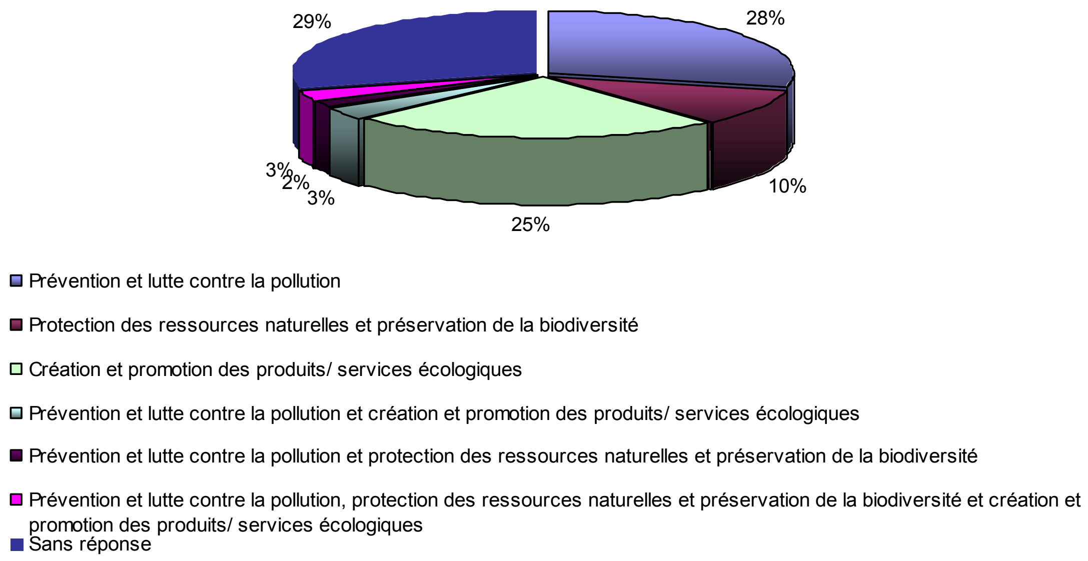
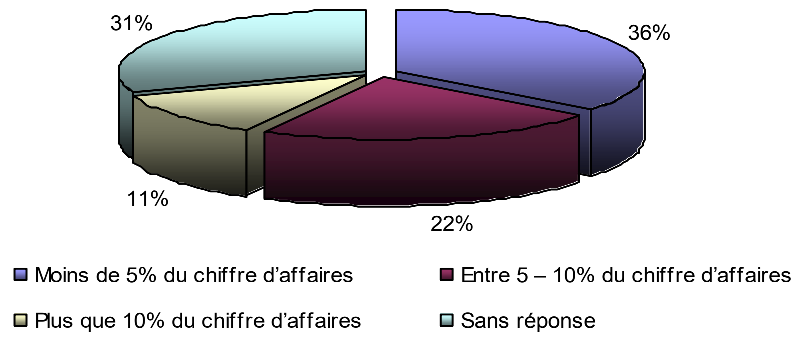


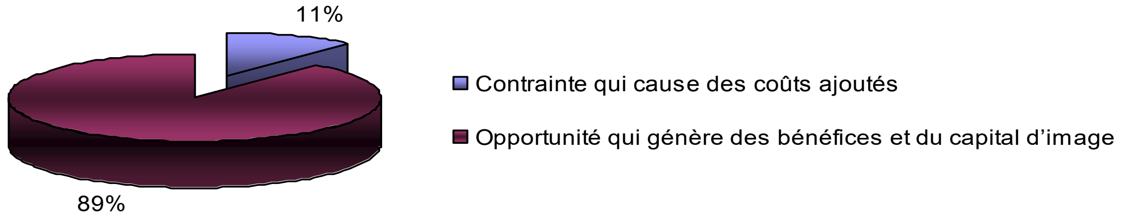
Figure 3: Le poids des investissements dans l’environnement sur le chiffre d’affaires



La question 4 - Vous percevez les investissements dans le domaine de la protection de l’environnement: **a.** comme une contrainte pour l’entité, qui cause des coûts ajoutés; **b.** comme une opportunité pour l’entité, qui détermine des bénéfices et du capital d’image dans la relation avec les partenaires d’affaires, les consommateurs etc.

Dans la période étudiée, on a constaté que le poids majoritaire (29%) du total des sujets analysés est représenté par des micro entreprises qui n’ont fait aucun investissement concernant l’environnement (Figure 2), même si seulement 11% des PME investigués perçoivent les investissements dans l’environnement comme une contrainte qui cause des coûts ajoutés, et 89% des sujets analysés considèrent que ces investissements génèrent des bénéfices et du capital d’image dans la relation avec les partenaires d’affaires et les consommateurs (Figure 4).

Figure 4: La perception des investissements dans l’enrironnement

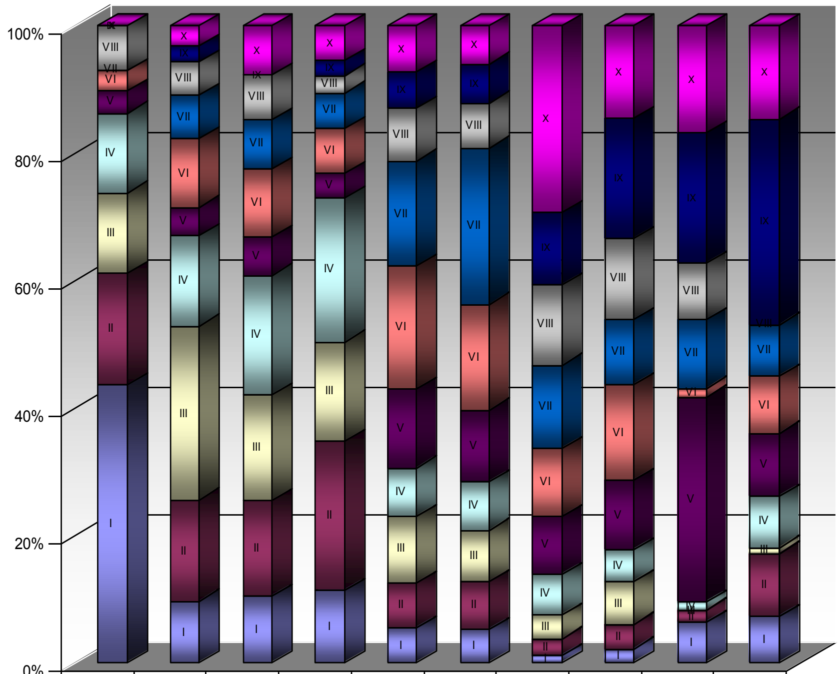


Les investissements en actions de prévention et lutte contre la pollution (la création de hales écologiques pour déposer les déchets et les substances dangereuses, l’organisation des systèmes de gestion des déchets électriques et électrotechniques, des installations de traitement des eaux usées, des installations d’épuration sèche des gazes de combustion et des substances polluantes etc.) ont été accomplis par 28% des sujets, des sociétés cotées et des moyennes entreprises. Tout aussi signifiant est le poids (25%) de ceux qui ont investi dans la création et la promotion des produits/services écologiques ou (10%) dans la protection des ressources naturelles et dans la préservation de la biodiversité (la réhabilitation écologique du sol et de l’eau phréatique en utilisant des technologies innovatrices etc.). Un pourcentage de 8% des sujets a investi en deux ou trois des activités mentionnées antérieurement (Figure 2). Pour 36% du total des entreprises, la valeur totale des investissements dans l’environnement ne dépasse pas 5% du chiffre d’affaires correspondant à la période analysée (Figure 3). L’hypothèse no. 2 est en partie confirmée par les résultats antérieurs. Même si, en réalité, le coefficient de pondération des sujets qui n’ont pas fait des investissements dans l’environnement est signifiant (29%), et dans le cas de ceux qui ont fait de tels investissements (surtout les PME) leur coefficient représente moins de 5% du chiffre d’affaires, on peut observer des changements de mentalité et de perception vis-à-vis des principes de la préservation de l’environnement. Ainsi, les investissements dans l’environnement sont perçus par la plupart des répondants comme une opportunité, une occasion de démontrer un comportement responsable en matière d’environnement, qui peut se matérialiser en profit et en capital d’image, et non pas seulement en coûts et risques.

Hypothèse 3: *Les initiatve sociales représentent un des outils le moins souvent communiqué et utilisé par les dirigeants pour suivre et anticiper la performance, par raport aux indicateurs comptables et financiers traditionnels* est testée par la question 8.

La question 8 - Etablissez l'ordre (en employant des chiffres entre 1 et 10, 1= le plus important) de l'importance des éléments suivants dans la perspective de votre entité: **a.** le profit; **b.** les relations contractuelles équilibrées avec les terces (clients, fournisseurs, créditeurs etc.); **c.** le chiffre d'affaires; **d.** la rentabilité; **e.** la cote du marché; **f.** la liquidité et la solvabilité; **g.** les initiatives sociales (la création des emplois et la sécurité du travail, la protection de l'environnement, l'implication dans la vie de la communauté, des sponsorisations etc.); **h.** l'image de l'entité; **i.** le cours de l'action; **j.** le prestige de l'équipe directrice.

Figure 5: Hérarchisation des critéres de performance



	Profit	Relations avec les tierces	Chiffre d'affaires	Rentabilité	Cote du marché	Liquidité/ solvabilité	Initiatives sociales	L'image de l'entité	Cours boursièr	Le prestige
X	0	3.14	7	5.42	7.32	6	29.29	14.61	16.82	10.4
IX	0	2.55	0	2.4	5.6	6.2	11.45	18.78	20.42	22.94
VIII	7.1	5.25	6.33	2.75	8.35	7	12.59	12.74	8.91	0
VII	0	6.73	6.95	5.55	16.39	24.54	13.1	10.19	10.82	5.73
VI	3.18	10.82	9.7	7	19.32	16.58	10.64	15.08	1.3	6.38
V	3.51	4.46	5.46	3.91	12.41	11.3	9	10.82	32.13	7
IV	12.46	14.17	16.95	22.66	7.65	7.6	6.38	5.1	1.3	5.73
III	12.55	27.31	15	15.57	10.46	8	3.82	6.66	0	0.63
II	17.51	15.9	13.44	23.35	7.03	7.48	2.55	3.82	1.92	7
I	43.69	9.67	9.51	11.39	5.47	5.3	1.18	2.2	6.38	5.21

Conformément aux données de la Figure 5, on observe que, pour l'évaluation des performances et des perspectives des entreprises, les critères les plus pertinents (la première place en tant qu'importance) sont considérés : le profit (par 43,69% des sujets questionnés); la rentabilité (par 11,39%); les relations contractuelles équilibrées avec les terces (par 9,67%); le chiffre d'affaires (par 9,51%); le cours boursier (par 6,38%); la cote du marché (par 5,47%); la liquidité et la solvabilité (5,30%); le prestige de l'équipe directrice (par 5,21%); l'image de l'entité (par 2,20%); les initiatives sociales (seulement par 1,18%). On peut aussi remarquer que 29,29% des sujets interviewés placent le critère « initiatives sociales » sur la 10è position de dix. Ces données confirment la forte stabilité des critères comptables et financiers traditionnels, même dans l'environnement dynamique du 21è siècle, pour apprécier les performances de l'entreprise. Les initiatives sociales, à l'avis des répondants, représentent l'un des outils les moins souvent communiqués et utilisés par les dirigeants, notamment ceux des PME, pour suivre et anticiper la performance.

3.3. Discussions

Les sociétés roumaines cotées à la Bourse, qui entrent en compétition sur le marché international, sont les plus impliquées dans la vie de la communauté, ayant un comportement proactif dans cette direction. Leurs stratégies incluent systématiquement et volontairement des objectifs, ressources et procédures d'amélioration continue de l'impact social et environnemental de leur activité. Ces sociétés sont les premières à avoir opéré des changements dans les structures organisatrices pour localiser les responsabilités de la fonction environnementale et les attributions des *éco dirigeants* qui ont la mission de formuler des polices environnementales et sociales, de contrôler l'évolution des normatives et le respect des standards de l'environnement, de communiquer avec les parties prenantes au propos de leur comportement socialement responsable. Pour la plupart, mais surtout les sociétés qui opèrent dans le domaine de l'énergie, de la chimie, de la pharmacie etc., ont implémenté et certifié, conformément aux standards internationaux spécifiques (ISO 9001; ISO 14001; OHSAS 18001; SA 8000; AA 1000; ISO 26000), des systèmes intégrés de gestion «Qualité – Environnement – Sécurité » et communiquent avec les parties prenantes au propos de leur performance pour obtenir une réponse, en termes de profit et d'image, aux investissements faits dans ce but.

Les petites et moyennes entreprises manifestent, avec des petites exceptions, un comportement passif vis-à-vis des aspects environnementaux. Elles sont soit indifférentes à la gestion de leurs propres prestations environnementales et perçoivent l'environnement comme une contrainte qui cause des coûts ajoutés, qu'elles ne veulent pas transformer dans une opportunité, dans une logique où le profit est le seul but de ces entreprises, soit elles adoptent un comportement responsable minimum, imposé par la loi. Ces entreprises n'ont pas des initiatives volontaires constantes dans la direction de la protection de l'environnement qu'après être devenues les sujets de certains incidents ou catastrophes environnementales. Elles sont disposées à faire un minimum de dépenses en actions sociales ou environnementales, pour s'assurer la fonctionnalité. Dans la plupart des cas, les entreprises qui ont des initiatives sur le plan social et environnemental, le font d'une manière occasionnelle, au niveau local et indépendamment de la stratégie de l'entreprise. Ainsi, leurs actions répondent d'une manière essentielle à l'activité de nature philantropique. Une conclusion positive est le fait que l'évolution du domaine des affaires en Roumanie, l'ouverture vers l'économie globale et l'augmentation du rôle de la société civile tendent à changer la mentalité et la perception des individus qui deviennent plus attentifs vis-à-vis des principes de la préservation de l'environnement, des droits de l'homme et de la sécurité des conditions de travail.

La philosophie du développement durable et implicitement de la responsabilité sociale n'est pas contraire au concept d'efficacité. Elle ne se propose pas d'éliminer le profit de l'équation de la dynamique de l'économie et de la société, mais simplement de démontrer que le profit ne peut pas être envisagé par lui-même. Le profit reste la condition nécessaire pour assurer le développement durable, mais la société désire de la transparence en ce qui concerne sa manière de se former. En plus, le concept de responsabilité sociale de l'entreprise dépasse l'acception de philanthropie. La contribution de l'entreprise au développement durable ne consiste pas en ce qu'une entreprise responsable fait avec 1% de son profit, mais plutôt en la manière qu'elle gagne 99% de son profit. Si, par exemple, le personnel de l'entreprise travaille dans un milieu nuisible, si la production polue l'écosystème, lorsque la communication de l'entrprise avec l'extérieur est peu transparente et les produits ne respectent pas les normes d'hygiène et sécurité, l'entreprise aura des difficultés dans l'accomplissement des objectifs de competitivité et allouant 1% son profit aux individus auxquels elle a causé ces effets adverses, son impact dans le sens de la responsabilité sociale est négatif. Seulement par « l'internalisation » des coûts externes, soutiennent les spécialistes, il sera possible de passer au niveau de la production optimale d'un marché orienté vers le profit (le coût marginal égal avec le revenu marginal) à un niveau de la production socialement optimale. Cette stratégie d'affaires, plus inclusive, par laquelle on essaie de gérer les externalités négatives par leur internalisation, est un instrument de changement interne et externe de l'entreprise, qui représentent les prémices des opportunités innovatrices ultérieures, de la consolidation des relations de l'entreprise avec les parties prenantes et pour trouver des solutions à des problèmes auxquels se confronte la société dans l'époque de la globalisation.

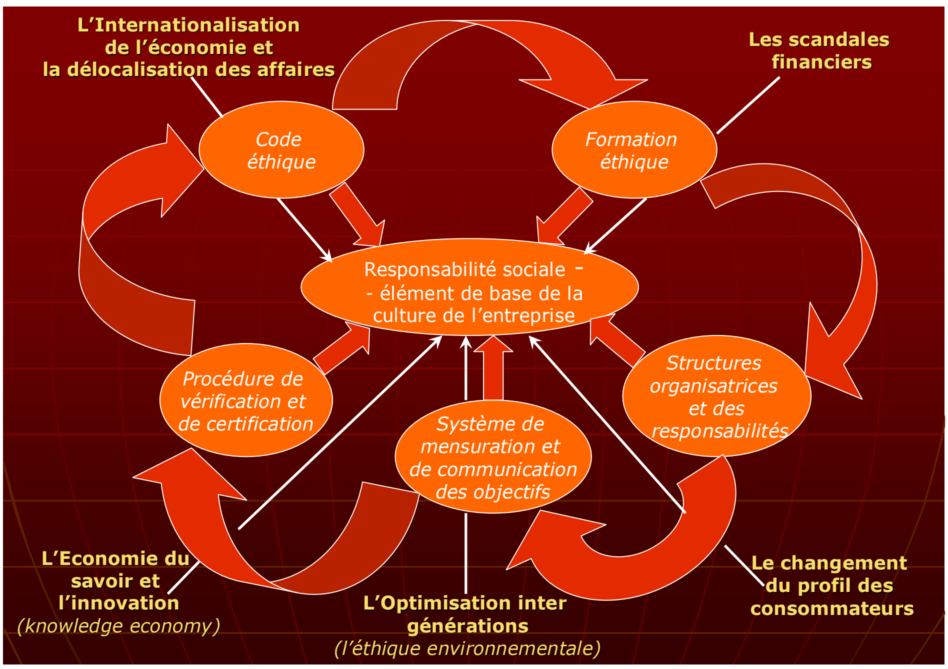
4. Conclusions

L'entreprise est le résultat d'un « pacte de non-agression » entre les sujets qui participent à sa vie économique. La conciliation et la satisfaction durables des intérêts antithétiques de tous les « acteurs sociaux » liés à l'activité de l'entreprise représentent une condition pour son développement et l'expression de la responsabilisation sociale de l'entreprise durable.

La responsabilité sociale de l'entreprise peut être comprise comme une norme sociale du comportement de l'entreprise, par laquelle celle-ci acquiert sa légitimité et l'accord de la société civile quant à ses actions. L'attention majeure accordée aux problèmes sociaux et leur inclusion dans les principes de gestion de l'entreprise représentent la prise en conscience de la nécessité d'adopter un modèle de gestion qui tienne compte des exigences des interlocuteurs sociaux. La responsabilité sociale de l'entreprise ne représente pas seulement une contrainte pour que l'entreprise actionne d'une manière correcte du point de vue éthique, social et écologique, mais elle a aussi la connotation d'un objectif stratégique pour l'entreprise, en mesure de garantir un consensus en ce qui concerne les actions qu'elle entreprend dans son environnement économique.

La responsabilité sociale représente un élément de base de la culture de l'entreprise. La formation d'une culture pro-responsabilité sociale dans l'entreprise ne peut survenir que comme suite à l'existence d'un *leadership* avec une vision éthique, en mesure de définir l'identité sociale de l'entreprise conformément aux critères d'équité entre les attentes multiples des parties prenantes. Cela suppose : la fondation d'un code éthique pour définir le système de valeurs qui guide le comportement de l'entreprise dans sa relation avec les interlocuteurs sociaux; la formation éthique, un processus de nature culturelle, par lequel les acteurs sociaux se familiarisent avec et acceptent l'application des normes et des principes sociaux dans le code éthique, dans le cadre de leurs actions; des structures organisatrices et des responsabilités précises pour l'alignement de la stratégie et des objectifs de l'entreprise aux valeurs et aux principes éthiques acceptés; un système de mensuration et de communication du degré d'atteinte des objectifs de la responsabilité sociale, pour obtenir une réponse, en termes de profit et d'image, aux investissements faits dans ce but, de la part des parties prenantes; une procédure de vérification et de certification de la conformité du système de gestion de l'entreprise aux normes de la responsabilité sociale de l'entreprise, conformément aux standards internationaux spécifiques (Figure 6.).

Figure 6: La responsabilité sociale un élément de la culture de l'entreprise



A l'affirmation de cette *nouvelle culture de l'entreprise* contribue aussi la société civile, dont le rôle a augmenté dans tous les pays, parce qu'au niveau de l'opinion publique les citoyens croient plutôt en les informations provenues des organisations non-profit qu'en celles provenues des entreprises ou des gouvernements.

Les compagnies roumaines doivent adapter leur stratégie d'affaires en rapport avec ces changements, réévaluer la durabilité des actions qu'elles entreprennent, en mesurant et en gérant leur propre impact social, environnemental et économique. L'étude empirique qui a analysé, d'une manière très générale, l'attitude et la réceptivité des entreprises roumaines concernant la dimension environnementale de la responsabilité sociale conduit au tracement de quelques directions d'action qui ont comme but l'amélioration future de la situation constatée, comme par exemple :

- *le transfert des bonnes pratiques en matière de responsabilité sociale* à partir des compagnies multinationales dans le top des 10 qui ont investi dans la Roumanie dans cette direction (Petrom S.A., Vodafone Roumanie S.A., Holcim Roumanie S.A., A&D Pharma, Raiffeisen Bank Roumanie, S. C. UniCredit Bank, Agricola Internaţional S.A. Bacău, S.C. Alexandrion Grup Roumanie, Lukoil Roumanie, Orange Roumanie) vers les entreprises non impliquées dans ces actions, en sachant que la communication fait partie d'une approche plus moderne de la responsabilité sociale;
- *la promotion des partenariats entre les administrations publiques – les compagnies privées - les organisations non gouvernementales* en mesure à contribuer à une meilleure identification et solution des problèmes sociaux et environnementaux pour lesquels des solutions n'ont pas encore été trouvées;
- *le développement des partenariats entre les entreprises et les universités (instituts de recherche)* pour le développement du savoir, l'implémentation des principes et instruments nécessaires pour la formation de la culture et du comportement responsable du point de vue social, conformément aux prévisions de l'Agenda de Lisbonne;
- *l'encouragement des PME à approcher leur gestion de manière plus intégrée,* afin d'y inclure les aspects sociaux et environnementaux.

Mentionnons que la taille de l'échantillon et le fait que c'est une étude à caractère exploratoire plutôt que confirmative limitent les possibilités de généralisation. Le sondage amélioré, de sorte qu'il permette la collecte des informations relatives à toutes les facettes de la responsabilité sociale (relations avec les parties prenantes, le préoccupation pour la santé, la sécurité et la motivation du personnel, la préservation de l'environnement), pourrait représenter un instrument important de diagnostic des performances de la responsabilité sociale des entreprises roumaines en comparaison avec l'expérience des pays européens, pour une future piste de recherche. L'administration du sondage à une population plus grande et plus diversifiée permettra non seulement la validation ou l'infirmation statistique des résultats, mais aussi une amélioration considérable des résultats de la recherche.

5. Bibliographie:

[1] Benedetti, F., Foti, Al., Guardavilla, A., (2007), *I sistemi di gestione per la sicurezza, la salute e l'ambiente*, Milano, Edizioni Professionali.

[2] Bergeron, H., (2002), *La gestion stratégique et les mesures de la performances non financière des PME, 6°* Congrès International Francophone sur la PME, HEC Montréal, Octobre 2002.

[3] Bianchini, M., (2004), *Perché alle organizzazioni del XXI secolo si chiede di essere Socialmente Responsabili? In che modo possono realizzare un comportamento etico e come possono comunicarlo?*, 2004, http://www.itconsult.it/knowledge/articoli/pdf/itc_bianchini.

[4] Cămăşoiu, C., Popescu, S., Cernat, R., Pană, C., (2003), *Identificarea decalajelor dintre România şi ţările membre UE în privinţa instrumentelor economice de protecţie a mediului. Studiu de caz: protecţia apei*, in: Seria *Probleme Economice, Vol.74/*CIDE, nr. 39, 2003.

[5] Chauzal, Ch., Murer, A., (2007), *Réflexion sur la contribution de la R.S.E. à la performance de l'entreprise: l'approche par les congés de solidarité internationale*, Brest Bretagne, Colloque Internationale: *À la recherche de la performance: Un concept de gestionnaire essentiel à l'entreprise*", 1er et 2 février 2007.

[6] Commoner, B., (2003), *Una valutazione del progresso ambientale: la ragione del fallimento*, Vol. *Economia dell'ambiente e bioeconomia*, a cura di Molesti R., Milano, FrancoAngeli.

[7] Del Bello, A. Gasperini, A., (2006), *Il valore del capitale intellettuale*, Milano, IPSOA Editore.

[8] Feleagă, N., (2006), *Criza financiară la cumpăna dintre secolele XX şi XXI şi guvernanţa întreprinderii, Revista Economie Teoretică şi Aplicată*, Nr. 9, p. 61 – 69.

[9] Freeman, R.E., Werhane, P., (1999), *Business Ethics: The State of the Art*, Journal of Management Review, Vol. 1, n° 1, p. 1 – 16.

[10] Friedman, M., (1962), *Capitalism and Freedom*, Chicago University Press.

[11] Germain, Ch., Gates, S., (2007), *Le niveau de développement des indicateurs de responsabilité sociale dans les outils de pilotage de contrôle de gestion: une analyse des pratiques des entreprises,* 28eme Congres de l'Association Francophone de Comptabilité: *Comptabilité et environnement*, Poitiers, 23-25 mai 2007.

[12] Hulot, N., (2006), *Pour un pacte ecologique*, Paris, Edition Calmann-Lévy.

[13] Janicot, L., (2007), *Les systèmes d'indicateurs de performance environnementale (IPE), entre communication et contrôle*, *La Revue Comptabilité Contrôle Audit*, Tome 13, p. 47 – 67.

[14] Longo, E., Bartolomeo, M., (1998), *Ambiente, comunicazioni, società. Negoziare il futuro sostenibile*, Milano, Il Sole 24 Ore.

[15] Marchetti, S., (2006), *L'informativa socio-ambientale nel sistema aziendale. Strumenti Comunicazionali*, Bari, Cacucci Editore.

[16] Mironiuc, M., (2005), *Analiza sistemelor de management ambiental. Regulamerntul EMAS şi Normele din familia ISO 14000*, Iaşi, Editura Politehnium.

[17] Mironiuc, M., Bedrule, Grigoruţă M. V., (2007), *La performance dans la communication economico-financiere de l'entite. Potentiel cognitif et limites*, Bretagne Brest, Colloque Internationale: *À la recherche de la performance: Un concept de gestionnaire essentiel à l'entreprise*, 1er et 2 février 2007.

[18] Molesti, R., (2003), *Economia dell'ambiente e bioeconomia*, Milano, FrancoAngeli.

[19] Pearce, D.W., Turner, R.K., (1991), *Economia delle risorse naturali e dell'ambiente*, Bologna, Il Mulino.

[20] Perez, R., (2005), *Pour un approche socialement responsable du management*, Paris, Economica.

[21] Pohoaţă, I., (2003), *Fiosofia economică şi politica dezvoltării durabile*, Bucureşti, Edition Economică.

[22] Poli, C., (1994), *Etica ambientali, teoria e pratica*, Milano, Ed. Guerini Studio.

[23] Reynaud, E., (2006), *Le développement durable au coeur de l'entreprise*, Ed. Dunod, Paris.

[24] Sacconi, L., (2003), *Dobbiamo chiedere alle imprese di essere socialmente responsabili? E se sì, come?*, CELE.

[25] Stancu, A., Orzan, M., (2007), *Responsabilitatea socială a companiilor româneşti – un pas pentru dezvoltarea durabilă*, Revista de Marketing Online, Vol. 1., Nr. 2, p. 1-13, 2006, http://ideas.repec.org/p/ase/wpaper/1001.html.

[26] Turner, K.R., Pearce, W.D., Bateman, I., (2003), *Economia ambientale*, Bologna, Il Mulino.

[27] Vercelli, Al., Borghesi, S., (2005), *La sostenibilità dello sviluppo globale*, Roma, Carocci Editore.

[28] *** http://www. *europa.eu.int/comm/environment/eussd*, 8.05. 2004.

A NEW METHOD OF ROBUST LINEAR REGRESSION ANALYSIS: SOME MONTE CARLO EXPERIMENTS

Sudhanshu Kumar **MISHRA**
Department of Economics, North-Eastern Hill University,
Shillong, Meghalaya, **INDIA**
mishrasknehu@yahoo.com

Abstract:

This paper has elaborated upon the deleterious effects of outliers and corruption of dataset on estimation of linear regression coefficients by the Ordinary Least Squares method. Motivated to ameliorate the estimation procedure, it introduces the robust regression estimators based on Campbell's robust covariance estimation method. It investigates into two possibilities: first, when the weights are obtained strictly as suggested by Campbell and secondly, when weights are assigned in view of the Hampel's median absolute deviation measure of dispersion. Both types of weights are obtained iteratively and using those weights, two different types of weighted least squares procedures have been proposed. These procedures are applied to detect outliers in and estimate regression coefficients from some widely used datasets such as stackloss, water salinity, Hawkins-Bradu-Kass, Hertzsprung-Russell Star and pilot-point datasets. It has been observed that Campbell-II in particular detects the outlier data points quite well. Subsequently, some Monte Carlo experiments have been carried out to assess the properties of these estimators whose findings indicate that for larger number and size of outliers, the Campbell-II procedure outperforms the Campbell-I procedure. Unless perturbations introduced to the dataset are numerous and very large in magnitude, the estimated coefficients are also nearly unbiased.

Keywords: robust regression, Campbell's robust covariance, outliers, Monte Carlo experiment, median absolute deviation

JEL classification: C13, C14, C63, C15, C01

1. Introduction

The outliers in a dataset are the points in a minority that are highly unlikely to belong to the population from which the other points (i.e. inliers), which are in a majority, have been drawn. Alternatively, the outliers exhibit a pattern or characteristics that are alien or non-conformal to those of the inliers. Stated differently, if a majority of data points, $p_i \in p$, lie in a range (a, b), then a minority of data points, $q_j \in q$, far exterior to (a, b), are outliers in the data set $D \approx p \cup q$. The said range that divides D into p and q is often fuzzy since the definition of 'far exterior' cannot be exact. The points in the 'near exterior', which belong neither to p nor to q are in the indeterminate zone and to consider them the outliers or the inliers often needs some criterion, often ad hoc or presumptive in nature.

In any case, outliers in a data set pull the measures of central tendency towards themselves and also inflate the measures of dispersion leading to biased and inefficient estimators. The pulled measures of location and inflated measures of dispersion often lead to masking of the outliers. A single prominent outlier can mask other relatively less prominent outliers and thus may cause delusion and evade their detection by a cursory inspection.

2. Linear Regression Analysis

On many occasions we desire to explain changes in a dependent variable (Y) as a response to changes in (a single or multiple) explanatory variables (X) and we hypothesize that the relationship between Y and X is linear. That is to say that the data set is described as $Y = b_1X_1 + b_2X_2 + ... + b_mX_m$ or, in another sense, $Y = \frac{\partial Y}{\partial X_1}X_1 + \frac{\partial Y}{\partial X_2}X_2 + ... + \frac{\partial Y}{\partial X_m}X_m$. We obtain a data set $Y(n,1)$ and $X(n,m)$ such that $n \geq m$. This dataset may be presented as a system of n equations in m unknowns or, in matrix representation, $Y = Xb$. If $n > m$ and the equations are inconsistent among themselves, no b will exactly satisfy the relationship $Y = Xb$, but a residual, $e(n)$, will make up $Y = Xb + e$. From this, we have $X^{-g}Y = X^{-g}Xb + X^{-g}e$, where X^{-g} is the generalized inverse of X. Since $X^{-g} = (X'X)^{-1}X'$ such that $X^{-g}X = (X'X)^{-1}X'X = I$, we have $(X'X)^{-1}X'Y = b + (X'X)^{-1}X'e$. We assume X and e to be uncorrelated such that $X'e = 0$ whence we obtain $\hat{b} = (X'X)^{-1}X'Y$. This procedure of estimation of b is known as the method of (ordinary) least squares or the OLS.

The method of ordinary least squares is very powerful, but at the same time it is very sensitive to contamination in Y or X and the nature of e as well the relationship between X and e. As for the residuals (e), it is required that each e_i should have zero mean and constant (non-zero) standard deviation, or $\mathrm{E}(e_i) = 0$; $E(e_i^2) = \sigma^2 \neq 0$, where $\mathrm{E}(\square)$ is the (statistical) expectation of $(\square)$. It is also necessary that $\mathrm{E}(e_L e_T) = 0$, where e_L and e_T are leading and trailing points, which is relevant only if the data points obey some order such as one in the time series. Together, these requirements are summarized to state that $E(ee') = \sigma^2 I$. As to X and its relationship with e, it is necessary that $\mathrm{E}(X'e) = 0$. Normally, X should be fixed or non-stochastic. If these conditions are satisfied, the OLS provides the BLUE or best linear unbiased estimator (of the parameters, b). These requirements are collectively called as the Gauss-Markov conditions [Plackett, (1950); Theil, (1971)]. It may be noted that the OLS to be BLUE does not require e_i to be normally or even identically distributed.

Aitken (1935), who was perhaps the first statistician to present the method of the ordinary least squares in matrix notations, extended the OLS to his Generalized Least Squares (GLS) to take care of the cases when $\mathrm{E}(ee') = \Omega \neq \sigma^2 I$. The GLS-estimated b (to be denoted by b_{GLS}) is obtained as $b_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$. Since Ω (and hence Ω^{-1} too) is a symmetric positive definite matrix, we may factorize $\Omega^{-1} = \omega'\omega$, whence $b_{GLS} = [(\omega X)'(\omega X)]^{-1}(\omega X)'(\omega Y)$. In this sense, the GLS is a weighted least squares, where ω is the weight matrix. In the OLS we have $\omega = (1/\sigma)I$. Aitken showed that the GLS estimators are BLUE. In particular, when the off-diagonal elements of Ω are all zero, we have $\omega_{ii} = 1/\sigma_i$ for all $i = 1, n$ and $\omega_{ij} = 0$ for $i \neq j$; $i, j = 1, n$.

3. The Case of Contaminated Datasets

In spite of meeting all the conditions mentioned above, contamination of the dataset makes the OLS an unsatisfactory method of estimation. This fact can be demonstrated by a simple example.

Table 1. Generated Data Set to Demonstrate the Effect of Mutilation by Introduction of Outliers

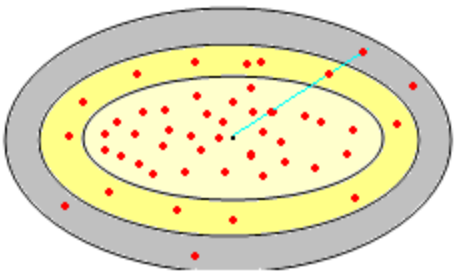
Original (Generated) Data Set								Mutilated Data Set							
Sl	Y	X_1	X_2	Sl	Y	X_1	X_2	Sl	Y	X_1	X_2	Sl	Y	X_1	X_2
1	176.1168	10.1683	10.3990	9	181.7555	10.9547	10.7604	1	176.1168	10.1683	10.3990	9	181.7555	**10.9547**	10.7604
2	170.4097	10.8740	10.0387	10	194.5983	11.5369	11.5459	2	170.4097	10.8740	10.0387	10	194.5983	11.5369	11.5459
3	222.4446	11.6551	13.2853	11	179.4799	12.8106	10.5921	3	222.4446	11.6551	13.2853	11	179.4799	12.8106	10.5921
4	209.2376	12.1879	12.4598	12	178.4198	12.0033	10.5327	4	209.2376	12.1879	12.4598	12	178.4198	12.0033	10.5327
5	193.3530	12.4701	11.4621	13	191.7982	11.4577	11.3721	5	193.3530	12.4701	11.4621	13	191.7982	11.4577	11.3721
6	192.5315	9.7968	11.4266	14	191.1635	11.3358	11.3349	6	192.5315	9.7968	11.4266	14	191.1635	11.3358	11.3349
7	164.5448	11.3915	9.6676	15	200.7671	11.1565	11.9338	7	164.5448	11.3915	9.6676	15	200.7671	11.1565	11.9338
8	164.8064	11.2341	11.6864	16	171.5413	11.5555	10.1067	8	164.8064	11.2341	**11.6864**	16	171.5413	11.5555	10.1067

The dataset presented in Table-1 (left panel) has been generated such that $Y = 8.7 + 0.1X_1 + 16X_2 + e$, where e is a very small disturbance. The ordinary least squares estimation of parameters from this data set gives $\hat{Y} = 8.74057 + 0.09872X_1 + 15.99787X_2$ which is very close to the generator equation. Next, we have mutilated $X_{8,2}$ and $X_{9,1}$ only slightly, which cannot possibly be detected by a mere eye inspection (right panel). Once again we apply the ordinary least squares estimation, which gives $\hat{Y} = 11.52752 + 0.98960X_1 + 14.65721X_2$. It may be noted that there is a tenfold increase in the magnitude of the coefficient associated with X_1. The value of R^2 has dropped down from 0.999998 to 0.760542. The moral of this story is clear: presence of outliers and corruption of only a few data points can sizably distort the estimated values of some or all parameters of the regression equation.

4. Detection of Contaminated or Outlier Data Points

If the contaminated or outlier data points can be detected, something can be done to eliminate them from the dataset or to abate their influence on the estimated regression coefficients. In particular, such data points can be assigned a relatively lower (even zero) weights vis-à-vis the inlier data points and a weighted least squares approach to estimation can be employed.

Figure 1: The Mahalanobis Distance



Mahalanobis (1936) defined his generalized distance, $d=[\{Y-\mathrm{E}(Y)\}'S^{-1}\{Y-\mathrm{E}(Y)\}]^{1/2}$, where the symbol S stands for the covariance matrix of Y. This distance is a measure of deviation of a (multivariate) data point from its center. If this distance is larger than a presumed value, the data point may be considered as an outlier. This measure is formally very similar to the device Ω used by Aitken in developing his Generalized Least Squares.

5. Campbell's Robust Covariance Matrix

Using the Mahalanobis distance as a measure of deviation from center, Campbell (1980) obtained a robust covariance matrix. Campbell's method is an iterative method that obtains the $m-$element vector of weighted (arithmetic) mean, $\bar{x}$, and weighted variance-covariance matrix, $S(m,m)$, in the following manner. Initially, all weights, $\omega_i; i=1,n$ are considered to be equal, $1/n$, and the sum of weights, $\sum_{i=1}^{n}\omega_i=1$. Further, we define $d_0=\sqrt{m}+b_1/\sqrt{2};\ b_1=2,\ b_2=1.25$. Then we obtain

$$\bar{x}=\sum_{i=1}^{n}\omega_i x_i \Big/ \sum_{i=1}^{n}\omega_i$$

$$S=\sum_{i=1}^{n}\omega_i^2(x_i-\bar{x})'(x_i-\bar{x}) \Big/ \left[\sum_{i=1}^{n}\omega_i^2-1\right]$$

$$d_i=\left\{(x_i-\bar{x})S^{-1}(x_i-\bar{x})'\right\}^{1/2}; i=1,n$$

$$\omega_i=\omega(d_i)/d_i; i=1,n: \omega(d_i)=d_i \ if\ d_i\le d_0\ else\ \omega(d_i)=d_0\exp[-0.5(d_i-d_0)^2/b_2^2].$$

If S is ill-conditioned for ordinary inversion, a generalized or the Moore-Penrose inverse [Theil, (1971)] of S or S^+ may be used for S^{-1} and if $d_i=0$ or $d_i\approx 0$ then $\omega_i=1$. We will call it the Campbell-I procedure to obtain a robust covariance matrix.

6. Use of Hampel's Median Absolute Deviation

Hampel *et al.* (1986) defined the median of absolute deviations (from median) as a measure of scale, $s_H^*(x_a)=\underset{i}{median}\,|\,x_{ia}-\underset{i}{median}(x_{ia})\,|$ and $s_H=s_H^*/0.6745$, which is a very robust measure of deviation. Using s_H, we may assign weights to different data points. If we heuristically assign the weight $\omega_i=1$ for $d_i-s_H(d)\le d_i<d_i+s_H(d)$, $\omega_i=(1/2)^2$ for $d_i-2s_H(d)\le d_i<d_i-s_H(d)$ as well as $d_i+2s_H(d)\ge d_i>d_i+s_H(d)$ and so on, and use Campbell's iterative method incorporating these weights, we may obtain a robust covariance matrix. Although not suggested so by Campbell (1980) himself, we will, however, obtain ω in this manner and call the resulting procedure as the Campbell-II method to obtain a robust covariance matrix.

7. Two Algorithms for Robust Regression Analysis

Let $Z=[Y\,|\,X]$. First, we obtain a robust covariance matrix $S=S(Z)$. In the process, we also obtain $\omega_i;\ i=1,n$. With ω we construct a matrix $W_{n,n}$ such that $w_{ij}=\omega_i$ for $i=j$ else $w_{ij}=0;\ i,j=1,n$. Then, using this weight matrix we obtain the robust regression estimator, $b_c=[(WX)'(WX)]^{-1}(WX)'(WY)$. In obtaining $S(Z)$ we may use the Campbell-I or the Campbell-II procedure and accordingly, we get two different b_c both of which are notably robust against data contamination and outliers.

8. Performance of Robust Regression Algorithms on Some Test Datasets

Many investigators in robust statistics [e.g. Andrews, (1974); Rupert and Carrol, (19809; Rousseeuw and Leroy, (1987); Kashyap and Maiyuran, (19939, etc.] have tested their methods on certain specific datasets that contain outliers. In particular, the datasets used by Rousseeuw and Leroy (1987) are available on http://www.uni-koeln.de/themen/Statistik/data/rousseeuw. Those datasets provide a good and widely accepted test bed for robust regression analysis. Among those the "stackloss datasets" [Brownlee, (1965)], water salinity dataset [Rupert and Carrol, (1980)], Hawkins-Bradu-Kass dataset [Hawkins *et al.*, (1984)], the Hertzsprung-Russell star dataset [analysed by Rousseeuw and Leroy, (1987)], and the Pilot-Point dataset [Daniel and Wood, (1971)] have been used here to test the performance of the presently proposed methods of robust regression. Other test datasets also could be used, but we consider that exercise unnecessary.

8.1. The Stackloss Dataset

The dataset describes the operation of a plant for the oxidation of ammonia to nitric acid. The stackloss (y) is a function of the rate (x_1), temperature (x_2) and acid concentration (x_3). The dataset has 21 observations or cases. The dataset is reproduced in the Table 2.

Table 2. Stackloss Dataset [Brownlee, (1965); Rousseeuw and Leroy, (1987)]

sl	y	x_1	x_2	x_3	sl	y	x_1	x_2	x_3	sl	y	x_1	x_2	x_3
1	42	80	27	89	8	20	62	24	93	15	8	50	18	89
2	37	80	27	88	9	15	58	23	87	16	7	50	18	86
3	37	75	25	90	10	14	58	18	80	17	8	50	19	72
4	28	62	24	87	11	14	58	18	89	18	8	50	19	79
5	18	62	22	87	12	13	58	17	88	19	9	50	20	80
6	18	62	23	87	13	11	58	18	82	20	15	56	20	82
7	19	62	24	93	14	12	58	19	93	21	15	70	20	91

It is widely acclaimed that the data points (1, 3, 4, 21) and possibly the point (2) are outliers. While the points (1, 3, 4, 21) are considered outliers, Kashyap and Maiyuran (1993) estimate the parameters as (-37.65, 0.80, 0.577, -0.067) of which the first is the y-intercept and the subsequent three are the coefficients associated with x_1, x_2 and x_3 respectively.

We applied Campbell-I robust estimator on the data, but it did not detect any outlier and therefore the estimated coefficients were the OLS estimates (-39.92, 0.716, 1.295, -0.152) only. However, Campbell-II detected the points (1, 2, 3, 4, 21) as clear outliers and the points (13, 17) as very mild outliers. The estimated coefficients were (-32.47, 0.852, 0.451, -0.132).

8.2. The Water Salinity Dataset

The water salinity (i.e., its salt concentration) dataset (Rupert&Carrol, 1980) comprises data on water salinity (y) as the dependent variable and lagged salinity (x_1), trend (x_2) and river discharge in North Carolina's Pamlico Sound (x_3) as the explanatory variables. The dataset has 28 points (Table 3).

In this dataset, Rousseeuw and Leroy's method detects the points (5, 16, 23, 24) as outliers whereas Rupert and Carrol's method detects (1, 11, 13, 15, 16, 17) as outliers. Kashyap and Maiyuran's method detects (5, 8, 15, 16, 17) as outliers for which the coefficients are (22.30, 0.724, -0.279, -0.786).

Table 3. Water Salinity Dataset [Rupert and Carrol, (1980); Rousseeuw and Leroy, (1987)]

sl	y	x_1	x_2	x_3	sl	y	x_1	x_2	x_3	sl	y	x_1	x_2	x_3	sl	y	x_1	x_2	x_3
1	7.6	8.2	4	23.005	8	8.2	8.3	5	21.862	15	10.4	13.3	0	23.927	22	14.1	13.6	5	21.005
2	7.7	7.6	5	23.873	9	13.2	10.1	0	22.274	16	10.5	10.4	1	33.443	23	13.5	15	0	25.865
3	4.3	4.6	0	26.417	10	12.6	13.2	1	23.830	17	7.7	10.5	2	24.859	24	11.5	13.5	1	26.290
4	5.9	4.3	1	24.868	11	10.4	12.6	2	25.144	18	9.5	7.7	3	22.686	25	12.0	11.5	2	22.932
5	5.0	5.9	2	29.895	12	10.8	10.4	3	22.430	19	12.0	10	0	21.789	26	13.0	12	3	21.313
6	6.5	5	3	24.200	13	13.1	10.8	4	21.785	20	12.6	12	1	22.041	27	14.1	13	4	20.769
7	8.3	6.5	4	23.215	14	12.3	13.1	5	22.380	21	13.6	12.1	4	21.033	28	15.1	14.1	5	21.393

The Campbell-I detects the points (5, 16) as outliers and yield the estimates of regression equation as (20.63 0.708 -0.202 -0.725). On the other hand, Campbell-II detects the points (5, 16) as clear outliers, points (23, 24) as severe outliers and points (9, 12, 15, 18, 19, 25) as very mild outliers. The estimated regression coefficients are (21.98, 0.722 , -0.276, -0.783).

8.3. Hawkins-Bradu-Kass Dataset

This dataset was artificially generated by Hawkins *et al.* (1984) and consists of 75 points of four variables, y, x_1, x_2 and x_3. It is widely held that the dataset has ten extreme outliers and four other points which obey the regression model, but are located away from other inliers [Kashyap and Maiyuran, (1993)].

Table 4. Hawkins-Bradu-Kass Dataset [Hawkins, Bradu & Kass, (1984); Rousseeuw and Leroy, (1987)]

sl	y	x_1	x_2	x_3	sl	y	x_1	x_2	x_3	sl	y	x_1	x_2	x_3
1	9.7	10.1	19.6	28.3	26	-0.8	0.9	3.3	2.5	51	0.7	2.3	1.5	0.4
2	10.1	9.5	20.5	28.9	27	-0.7	3.3	2.5	2.9	52	-0.5	3.3	0.6	1.2
3	10.3	10.7	20.2	31.0	28	0.3	1.8	0.8	2.0	53	0.7	0.3	0.4	3.3
4	9.5	9.9	21.5	31.7	29	0.3	1.2	0.9	0.8	54	0.7	1.1	3.0	0.3
5	10.0	10.3	21.1	31.1	30	-0.3	1.2	0.7	3.4	55	0.0	0.5	2.4	0.9
6	10.0	10.8	20.4	29.2	31	0.0	3.1	1.4	1.0	56	0.1	1.8	3.2	0.9
7	10.8	10.5	20.9	29.1	32	-0.4	0.5	2.4	0.3	57	0.7	1.8	0.7	0.7
8	10.3	9.9	19.6	28.8	33	-0.6	1.5	3.1	1.5	58	-0.1	2.4	3.4	1.5
9	9.6	9.7	20.7	31.0	34	-0.7	0.4	0.0	0.7	59	-0.3	1.6	2.1	3.0
10	9.9	9.3	19.7	30.3	35	0.3	3.1	2.4	3.0	60	-0.9	0.3	1.5	3.3
11	-0.2	11.0	24.0	35.0	36	-1.0	1.1	2.2	2.7	61	-0.3	0.4	3.4	3.0
12	-0.4	12.0	23.0	37.0	37	-0.6	0.1	3.0	2.6	62	0.6	0.9	0.1	0.3
13	0.7	12.0	26.0	34.0	38	0.9	1.5	1.2	0.2	63	-0.3	1.1	2.7	0.2
14	0.1	11.0	34.0	34.0	39	-0.7	2.1	0.0	1.2	64	-0.5	2.8	3.0	2.9
15	-0.4	3.4	2.9	2.1	40	-0.5	0.5	2.0	1.2	65	0.6	2.0	0.7	2.7
16	0.6	3.1	2.2	0.3	41	-0.1	3.4	1.6	2.9	66	-0.9	0.2	1.8	0.8
17	-0.2	0.0	1.6	0.2	42	-0.7	0.3	1.0	2.7	67	-0.7	1.6	2.0	1.2
18	0.0	2.3	1.6	2.0	43	0.6	0.1	3.3	0.9	68	0.6	0.1	0.0	1.1
19	0.1	0.8	2.9	1.6	44	-0.7	1.8	0.5	3.2	69	0.2	2.0	0.6	0.3
20	0.4	3.1	3.4	2.2	45	-0.5	1.9	0.1	0.6	70	0.7	1.0	2.2	2.9
21	0.9	2.6	2.2	1.9	46	-0.4	1.8	0.5	3.0	71	0.2	2.2	2.5	2.3
22	0.3	0.4	3.2	1.9	47	-0.9	3.0	0.1	0.8	72	-0.2	0.6	2.0	1.5
23	-0.8	2.0	2.3	0.8	48	0.1	3.1	1.6	3.0	73	0.4	0.3	1.7	2.2
24	0.7	1.3	2.3	0.5	49	0.9	3.1	2.5	1.9	74	-0.9	0.0	2.2	1.6
25	-0.3	1.0	0.0	0.4	50	-0.4	2.1	2.8	2.9	75	0.2	0.3	0.4	2.6

We have applied Campbell-I and Campbell-II methods to detect the outliers and estimate the coefficients of robust regression. Campbell-I detects the points (11, 12, 13, 14) as outliers and the estimated coefficients are (-0.828, 0.156, 0.106, 0.226). Campbell-II detects the points (1 through 14) as clear outliers, points (18, 53, 71, 72) as strong outliers and the points (19, 28, 29, 40, 47, 50, 55, 59, 67, 68) as very mild outliers. The estimated regression coefficients by the Campbell-II method are (-0.775, 0.1625, 0.1812, 0.06517). The OLS estimates of coefficients are (-0.38755, 0. 239185, -

0.334548, -0.383341). A comparison of the Campbell-II and the OLS estimates of regression coefficients show the damage done by the outliers.

8.4. The Hertzsprung-Russell Star Dataset

This data set was introduced by Rousseeuw and Leroy (1987). It has 47 points in two variables, logarithm of the light intensity of the star as the dependent variable (y) and logarithm of the effective temperature at the surface of the star as the explanatory variable (x_1). It has four very strong outliers, the so-called giant stars, represented by the points (11, 20, 30, 34).

Table 5. Hertzsprung-Russell Star Dataset [Rousseeuw and Leroy, (1987)]

sl	y	x_1	sl	y	x_1	sl	y	x_1	sl	y	x_1	sl	y	x_1
1	4.37	5.23	11	3.49	5.73	21	4.29	4.38	31	4.38	4.42	41	4.38	4.62
2	4.56	5.74	12	4.43	5.45	22	4.29	4.22	32	4.56	5.10	42	4.45	5.06
3	4.26	4.93	13	4.48	5.42	23	4.42	4.42	33	4.45	5.22	43	4.50	5.34
4	4.56	5.74	14	4.01	4.05	24	4.49	4.85	34	3.49	6.29	44	4.45	5.34
5	4.30	5.19	15	4.29	4.26	25	4.38	5.02	35	4.23	4.34	45	4.55	5.54
6	4.46	5.46	16	4.42	4.58	26	4.42	4.66	36	4.62	5.62	46	4.45	4.98
7	3.84	4.65	17	4.23	3.94	27	4.29	4.66	37	4.53	5.10	47	4.42	4.50
8	4.57	5.27	18	4.42	4.18	28	4.38	4.90	38	4.45	5.22			
9	4.26	5.57	19	4.23	4.18	29	4.22	4.39	39	4.53	5.18			
10	4.37	5.12	20	3.49	5.89	30	3.48	6.05	40	4.43	5.57			

The Campbell-I method detects the points (11, 20, 30, 34) as clear outliers, the point (7) as a strong outlier, and points (9, 14) as very mild outliers. The regression coefficients are (3.7789, 0.126). The Campbell-II detects the points (7, 9, 11, 14, 20, 30, 34) as clear outliers and the points (3, 5, 18, 25, 28, 33, 38, 41, 42, 43, 46) as very mild outliers. The estimated regression equations are (3.7415, 0.13688). Against this, the OLS estimates of the coefficients are (4.847, -0.1071). The OLS estimates indicate that light intensity decreases as the temperature increases, which is obviously misleading. The robust regression coefficient, however, is positive.

8.5. The Pilot-Plant Dataset

Daniel and Wood (1971) provide the dataset of 20 points in two variables, where the dependent variable (y) is the acid content determined by titration, and the explanatory variable (x_1) is the organic acid content determined by extraction and weighing.

Table 6. Pilot Point Dataset [Daniel and Wood, (1971); Rousseeuw and Leroy, (1987)]

sl	y	x_1	sl	y	x_1	sl	y	x_1	sl	y	x_1	sl	y	x_1
1	76	123	5	55	57	9	41	16	13	88	159	17	89	169
2	70	109	6	48	37	10	43	28	14	58	75	18	88	167
3	55	62	7	50	44	11	82	138	15	64	88	19	84	149
4	71	104	8	66	100	12	68	105	16	88	164	20	88	167

The Campbell-I method does not detect any outlier in this data and hence the estimated regression coefficients, (35.4583, 0.3216) are the OLS estimates. However, the Campbell-II method detects a single very strong outlier point (11), four strong outlier points (4, 10, 13, 15) and three very mild outlier points (2, 8, 14). None of the points is a clear outlier. The estimated regression coefficients are (36.190, 0.3137).

These tests indicate that in detecting the outliers (and yielding the estimates of robust regression coefficients), the Campbell-I method is rather blunt and the Campbell-II is very sensitive. Where the outliers are not much deviant from the center, the Campbell-I fails to detect them. But Campbell-II detects very mild outliers too, occasionally signaling false positive.

9. Some Monte Carlo Experiments

We generated artificially a forty points 'base data' on three variables (x_1, x_2 and x_3), obtained $y = 80 - 16x_1 + 12x_2 - 2x_3$, and add a very small error to it to meet the requirements of regression analysis.

We present the 'base data' in Table 7. These data have no outliers and the OLS regression coefficients are (80.0071, -16.0001, 12.0001, -1.9998).

Table 7. Generated (Artificial) Base Dataset for Introducing Outliers in Monte Carlo Experiments

sl	y	x_1	x_2	x_3	sl	y	x_1	x_2	x_3	sl	y	x_1	x_2	x_3
1	325.62554	5.48761	28.73814	5.75294	15	-320.93381	38.27473	25.78603	49.00109	29	-216.70787	46.50270	44.03694	40.58253
2	216.23692	4.39939	17.35204	0.79249	16	380.45138	-4.53886	18.59881	-2.30804	30	65.97138	13.22553	21.07927	27.69959
3	61.22945	23.83300	35.86611	33.93566	17	240.77406	3.03338	23.03262	33.57479	31	100.21909	17.65419	24.29101	-5.58615
4	-503.35919	41.23092	11.65128	31.77223	18	33.14312	25.33601	31.93026	12.38928	32	76.79387	21.43170	30.03208	10.35714
5	235.15369	16.11028	37.32482	17.48946	19	-230.16532	33.78115	26.40625	43.28304	33	-372.67139	43.93101	28.55601	46.25669
6	-25.00594	22.25174	25.10806	25.09546	20	226.89812	7.18313	22.76687	5.72171	34	522.41150	-4.43900	32.89357	11.66127
7	176.81192	9.90141	25.04717	22.64996	21	115.94871	0.69938	9.38464	32.74268	35	280.68193	-6.29805	7.42427	-5.37605
8	-233.23841	19.83920	4.29028	23.65183	22	319.34732	12.12583	40.15470	24.25351	36	-398.03503	32.25189	7.91189	28.43924
9	-73.33875	8.76262	3.63445	28.40374	23	272.84371	6.58704	28.01780	19.00688	37	-373.70179	39.99472	20.60245	30.50913
10	154.46532	1.84157	11.41580	16.51787	24	-23.69083	17.15236	19.89292	33.96668	38	-270.06547	24.93147	4.11388	0.26708
11	-55.74443	30.61515	33.49451	23.86843	25	-375.92013	34.26759	10.90453	19.25040	39	195.19782	15.80573	33.94128	19.57595
12	-396.83923	30.81785	5.96886	27.70361	26	-78.48044	33.31807	32.40769	7.14970	40	-82.85239	27.95279	27.01828	19.91053
13	-337.50365	40.83092	21.78991	12.82849	27	-405.38187	32.61658	7.07663	24.22386	$y = 80 - 16x_1 + 12x_2 - 2x_3$				
14	-9.86483	29.19798	35.25464	22.88996	28	425.10924	11.33292	48.91941	30.30439					

9.1. Experiment-1

We add one quantum of a random size between (-10, -5) and (5, 10) to equi-probably randomly chosen point of every variable (including y). We do this exercise 200 times and find mean, $\bar{b}$, standard deviation, $s(b)$, and root-mean-square, Rms, for each coefficient (having 200 replicates). Estimation is done by Campbell-I and Campbell-II methods. Then we change the number of perturbation quanta to be made to each variable to 2, 5 and 10 keeping other parameters of the experiment constant. The results are presented in Table 8.

Table 8. Results of the Monte Carlo Experiments for Perturbations between (-10, -5) and (5, 10)

NO	EM	$\bar{b}_0$	$\bar{b}_1$	$\bar{b}_2$	$\bar{b}_3$	$s(b_0)$	$s(b_1)$	$s(b_2)$	$s(b_3)$	Rms_0	Rms_1	Rms_2	Rms_3
1	C1	79.9653	-15.9975	12.0000	-2.0000	0.4637	0.0146	0.0171	0.0147	0.4650	0.0148	0.0171	0.0147
	C2	79.9484	-15.9982	12.0000	-1.9987	1.0798	0.0282	0.0388	0.0257	1.0811	0.0283	0.0388	0.0258
2	C1	80.0780	-15.9944	11.9882	-1.9991	1.9785	0.0395	0.0825	0.0452	1.9800	0.0399	0.0833	0.0452
	C2	79.8596	-15.9927	11.9992	-1.9993	1.4753	0.0426	0.0490	0.0405	1.4820	0.0432	0.0490	0.0405
5	C1	83.8509	-15.7211	11.6521	-2.0359	16.6481	0.5297	0.5722	0.5203	17.0877	0.5986	0.6697	0.5216
	C2	80.7855	-15.9914	11.9688	-2.0015	4.6661	0.1116	0.1441	0.1247	4.7317	0.1120	0.1474	0.1247
10	C1	83.1449	-15.2665	11.3806	-2.1560	24.2318	0.7452	0.8149	0.7094	24.4350	1.0456	1.0236	0.7263
	C2	83.4905	-15.7827	11.6428	-2.0122	28.4410	0.8036	0.8973	0.8302	28.6544	0.8324	0.9658	0.8303

Note: NO=No. of perturbations per variable; EM= Estimation Method; C1=Campbell-I method; C2=Campbell-II method.

9.2. Experiment-2

Next, we repeat the experiment with change in the size of perturbation quanta, but keeping everything else as elaborated in Experiment 1. The perturbation quanta now lie in a larger range of (-25, -20) and (20, 25). The results are presented in Table 9.

Table 9. Results of the Monte Carlo Experiments for Perturbations between (-25, -20) and (20, 25)

NO	EM	$\bar{b}_0$	$\bar{b}_1$	$\bar{b}_2$	$\bar{b}_3$	$s(b_0)$	$s(b_1)$	$s(b_2)$	$s(b_3)$	Rms_0	Rms_1	Rms_2	Rms_3
1	C1	79.9745	-15.9756	11.9848	-2.0048	1.4511	0.0483	0.0595	0.0597	1.4513	0.0541	0.0615	0.0599
	C2	79.9635	-15.9800	11.9854	-2.0015	3.0999	0.0856	0.1173	0.0938	3.1001	0.0880	0.1182	0.0939
2	C1	80.3301	-15.9319	11.9346	-2.0095	5.9550	0.2491	0.1728	0.1467	5.9641	0.2583	0.1847	0.1470
	C2	79.9951	-15.9494	11.9633	-2.0092	4.8627	0.1564	0.1658	0.1419	4.8627	0.1644	0.1698	0.1422
5	C1	102.2277	-13.3051	9.1346	-2.4067	42.5440	1.3607	1.3525	1.3570	48.0007	3.0190	3.1686	1.4167
	C2	81.6860	-15.9585	11.7975	-1.8904	11.9805	0.3729	0.4931	0.4217	12.0986	0.3752	0.5330	0.4357
10	C1	100.8889	-10.9714	7.2610	-2.5784	53.1775	1.4913	1.6434	1.4639	57.1331	5.2451	5.0159	1.5740

	C2	108.7593	-13.4929	8.5766	-2.1912	75.5747	2.4112	2.3043	2.1935	80.8618	3.4784	4.1267	2.2018

Note: NO=No. of perturbations per variable; EM= Estimation Method; C1=Campbell-I method; C2=Campbell-II method.

9.3. Experiment-3

Once again we repeat the experiment with further changes in the size of perturbation quanta, but keeping everything else as elaborated in Experiment 1. The perturbation quanta now lie in a still larger range of (-100, -50) and (50, 100). The results are presented in Table 10.

Table 10. Results of the Monte Carlo Experiments for Perturbations between (-100, -50) and (50, 100)

NO	EM	$\bar{b}_0$	$\bar{b}_1$	$\bar{b}_2$	$\bar{b}_3$	$s(b_0)$	$s(b_1)$	$s(b_2)$	$s(b_3)$	Rms_0	Rms_1	Rms_2	Rms_3
1	C1	80.9852	-15.7714	11.8010	-2.0469	7.3542	0.2993	0.3634	0.3857	7.4198	0.3766	0.4144	0.3885
	C2	80.4523	-15.7762	11.8439	-2.0677	11.5777	0.3792	0.4672	0.4666	11.5865	0.4403	0.4926	0.4715
2	C1	84.0103	-15.5509	11.5696	-2.1508	16.6481	0.7617	0.7323	0.7673	17.1243	0.8843	0.8495	0.7820
	C2	82.3807	-15.5821	11.6483	-2.1339	18.3389	0.7976	0.7269	0.7900	18.4927	0.9004	0.8075	0.8012
5	C1	114.2855	-10.9362	5.8761	-1.7315	92.7163	5.2181	3.9374	1.8410	98.8524	7.2713	7.2804	1.8605
	C2	89.6809	-15.1442	11.0092	-2.1745	40.8676	1.6265	1.6630	1.3783	41.9986	1.8379	1.9357	1.3893
10	C1	53.8971	-3.7093	1.7535	-1.2798	65.2209	2.6154	1.7382	1.4089	70.2505	12.5658	10.3929	1.5823
	C2	97.7946	-8.0965	3.9146	-1.7399	116.3931	5.2697	3.4541	2.1912	117.7455	9.4992	8.7922	2.2065

Note: NO=No. of outliers per variable; EM= Estimation Method; C1=Campbell-I method; C2=Campbell-II method.

9.4. Observations

For small perturbations both Campbell-I and Campbell-II do perform well, but if the number of perturbations is smaller, the Campbell-I performs better. This edge of Campbell-I over Campbell-II is lost with an increase in the number of perturbations. Secondly, as the size as well as the number of perturbations increase, the robust estimators by both the methods tend to become biased as reflected in the increasing difference between $s(b)$ and Rms values. It may be noted that for unbiased ness $s(b) = Rms$. It has been empirically observed that ten perturbations per variable amount to corruption of about 35 percent points in the dataset. Further, considering the size/magnitude of independent variables (x_1, x_2 and x_3) that lie between (-10, 50), a perturbation lying between (-100, -50) or (50, 100) is quite large. Such perturbations can always induce biases in the estimated coefficients. As it is observed, when the perturbations per variable is up to five in number, Campbell-II produces very good results even when the size of perturbations is large.

10. Concluding Remarks

In this paper we have elaborated upon the deleterious effects of outliers and corruption of dataset on estimation of linear regression coefficients by the Ordinary Least Squares method. Motivated to ameliorate the estimation procedure, we have introduced the robust regression estimators based on Campbell's robust covariance estimation method. We have investigated into two possibilities: first, when the weights are obtained strictly as suggested by Campbell and secondly, when weights are assigned in view of the Hampel's median absolute deviation measure of dispersion. Both types of weights are obtained iteratively. Using these two types of weights, two different types of weighted least squares procedures have been proposed. These procedures are applied to detect outliers in and estimate regression coefficients from some widely used datasets such as stack loss, water salinity, Hawkins - Bradu-Kass, Hertzsprung-Russell Star and pilot-point datasets. It has been observed that Campbell-II in particular detects the outlier data points quite well (although occasionally signaling false positive too as very mild outliers). Subsequently, some Monte Carlo experiments have been carried out to assess the properties of these estimators. Findings of these experiments indicate that for larger number and size of outliers, the Campbell-II procedure outperforms the Campbell-I procedure. Unless perturbations introduced to the dataset are sizably numerous and very large in magnitude, the estimated coefficients by the Campbell-II method are also nearly unbiased.

11. References:

[1] Aitken, A.C., (1935), *On Least Squares and Linear Combinations of Observations*, in: *Proceedings of the Royal Society of Edinburgh*, 55: 42 – 48.

[2] Andrews, D.F., (1974), *A Robust Method for Multiple Linear Regression*, in: T*echnometrics,* 16: 523 – 531.

[3] Brownlee, K.A., (1965), *Statistical Theory and Methodology in Science and Engineering,* Wiley, New York.

[4] Campbell, N.A., (1980), *Robust Procedures in Multivariate Analysis I: Robust Covariance Estimation*, in: *Applied Statistics*, 29 (3): 231 – 237.

[5] Daniel, C. and Wood, F.S., (1971), *Fitting Equations to Data*, Wiley, New York.

[6] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A., (1986), *Robust Statistics:The Approach Based on Influence Functions,* Wiley, New York.

[7] Hawkins, D.M., Bradu, D., and Kass, G.V., (1984), *Location of Several Outliers in Multiple Regression Using Elemental Sets*, in: *Technomenics,* 26: 197 – 208.

[8] Kashyap, R.L., and Maiyuran, S., (1993), *Robust Regression and Outlier Set Estimation using Likelihood Reasoning*, in: *Electrical and Computer Engineering ECE Technical Reports*, TR-EE 93-8, Purdue University School of Electrical Engineering. http://docs.lib.purdue.edu/ecetr/33/

[9] Mahalanobis, P.C., (1936), *On the Generalized Distance in Statistics*, in: *Proceedings of the National Institute of Science of India,* 12: 49 – 55.

[10] Plackett, R.L., (1950), *Some Theorems in Least Squares*, in: *Biometrika,* 37: 149 – 157.

[11] Rousseeuw, P.J., and Leroy, A.M., (1987), *Robust Regression and Outlier Detection,* Wiley. New York.

[12] Rupert, D., and Carrol, R.J., (1980), *Trimmed Least Squares Estimation in the Linear Model*, in: *Journal of American Statistical Association, 75:* 828 – 838.

[13] Theil, H., (1971), *Principles of Econometrics*, Wiley, New York.

Note: A Fortran Computer Program for both of the proposed methods is available from the author on request. Contact: mishrasknehu@yahoo.com

A SHORT SURVEY ON CHAOTIC DYNAMICS IN SOLOW-TYPE GROWTH MODELS

Cesare **PALMISANI**
Deptartamento Matematica e Statistica
Federico II University, Naples, **ITALY**
cpalmisani@unina.it

Abstract:

In this paper we review some Solow-type growth models, framed is discrete time, which are able to generate complex dynamic behaviour. For these models – put forward by Day (1982, 1983); Böhm and Kaas (2000); and Commendatore (2005) – we show that crucial features which could determine the emergence of regular or irregular growth cycles are (i) if the average saving ratio is constant or not; and (ii) the curvature of production function, representing the degree of substitutability between labour and capital. The lower the degree of substitutability, the higher the likelihood of complex behaviour.

Keywords: logistic map, Li-York chaos, growth models, local stability, triangle stability.

JEL Classification: E25, E32, O40.

1. Introduction

The analysis of the fundamental issues in dynamical macroeconomics usually begins with the study of two (one-sector and one-dimensional) growth models: the Ramsey model [Ramsey, (1928)] and the Solow model [Solow, (1956)]. In the Ramsey model a representative consumer has an infinite life horizon and optimizes his/her utility. In the Solow model consumption is not optimal the representative agent saves a constant fraction of his income. In the next sections we will describe the Solow model and a few models which are very close to the Solow one and are able to generate chaotic dynamics. We note here that researches in several direction have spanned from the Solow model. For example, the Solow model inspired the works of Shinkay (1960), Meade (1961), Uzawa (1961,1963), Kurz (1963) and Srinivasan (1964) on two-sector growth models. Following this line of research, works about two-sector models appeared on the *Review of Economic Studies* in the 1960s [Drandakis (1963), Takajama (1963, 1965), Oniki-Uzawa (1965), Hahn (1965), Stiglitz (1967), among others].

2. The Solow growth model in discrete time

Following Hans-Walter Lorenz (1989) and Costas Aziariadis (1993), we will develop a discrete time variant of the growth model due to Solow (1956). We consider *a single good economy, i.e. an economy in which only one good is produced and consumed*. We assume that time t is discrete, that is t = 0, 1, 2, … . The symbols $Y_t, K_t, C_t, I_t, L_t, S_t$ indicate economywide aggregates respectively equal to *income, capital stock, consume, investment, labor force, saving at time* t. The capital stock K_0 and labor L_0 at time 0 are given. The constant s denotes both the average and the *marginal savings rates* and the constant n denotes the *growth rate of population*. We consider s and n as given exogenously. The map F: $(K_t,L_t) \rightarrow F(K_t,L_t)$ is *the production function*. We assume that:

1. $Y_t = C_t + I_t$: *for all time t = 0, 1, …, the economy is in equilibrium, i.e. the supply of income* Y_t *is equal to the demand composed of the quantity* C_t *of good to consume plus the stock* I_t *of capital to invest (closed economy like a Robinson Crusoe economy)*;

2. $I_t = K_{t+1}$: *investment at time t corresponds to all capital available to produce at time t+1 (working capital hypothesis)*;

3. $S_t = Y_t - C_t = s\,Y_t$ $(0 < s < 1)$: *saving is a share of income*;

4. $Y_t = F(K_t, L_t)$, i.e. *at time t all income is equal to the output obtained by the inputs capital and labor*;

5. $L_t = (1+n)^t L_0$ $(n > 0)$: *the labor force grows as a geometric progression at the rate* (n).

From (1.) and (3.) in a short run equilibrium $Y_t = C_t + S_t$ or $I_t = S_t$. Thus, applying (2.) and (3.), we have $K_{t+1} = sY_t$. Finally, from (4.) we obtain $K_{t+1} = s\,F(K_t, L_t)$.

From the latter expression, $K_{t+1}/L_{t+1} = s\,F(K_t,L_t)/L_{t+1}$.

If F is *linear-homogeneous* (or it tells that *F exhibits constant returns to scale)*, i.e.

$F(\lambda K, \lambda L) = \lambda F(K, L)$ (for all $\lambda > 0$),

then we have $K_{t+1}/L_{t+1} = s\,L_t F\left(K_t/L_t ,1\right)/L_t(1+n)$.

We set $k_t = K_t/L_t$ (*capital-labor ratio or capital per worker*) and $y_t = Y_t/L_t$.(*output-labor ratio or output per worker*). We derive in this way the production function in the intensive form: $y_t = f(k_t) = f\left(K_t/L_t ,1\right)$. Therefore we get the equation of accumulation *for the Solow model in discrete time with the working capital hypothesis*:

$k_{t+1} = s\,f(k_t)/(1+n)$ (2.1)

If we assume that *capital depreciates at the rate* $0 \le \delta \le 1$ (*fixed capital hypothesis*), the capital available at time t +1 corresponds to $K_{t+1} = K_t - \delta K_t + I_t$, from which

$K_{t+1} = s\,F(K_t, L_t) + (1 - \delta) K_t$.

As before we get the following time-map for capital accumulation

$k_{t+1}(1+n) = s\,f(k_t) + (1 - \delta) k_t$ (2.2) or $k_{t+1} = h(k_t)$ (2.3)

where $h(k_t) = \dfrac{1}{1+n}[s\,f(k_t) + (1 - \delta) k_t]$.

We notice that I_t is the *gross investment* while $K_{t+1} - K_t = I_t - \delta K_t$ is *the net investment*.

Costas Azariadis (1993, p. 4) tells us that *this model captures explicitly a simple idea that is missing in static formulations: there is a trade-off between consumption and investment or between current and future consumption. The implications of this ever-present competition for resources between today and tomorrow are central to macroeconomics and can be explored only in a dynamic framework. Time is clearly of the essence*.

For example, if we use the *Cobb-Douglas* production function $f(k_t) = Bk^{\beta}$ (B > 0, $0 < \beta < 1$, k ≥ 0) – in intensive form - the eq. (2.1) becomes $k_{t+1} = h(k_t) = \left(sBk_t^{\beta}\right)/(1+n)$ (2.4).

For all k ≥ 0 and B > 0, if we assume (β > 0) we deduce only that $h(k_t)$ is strictly monotonically increasing, instead we need to use both inequalities ((β > 0) and (β < 1) t show the concavity of $h(k_t)$.

As a matter of fact, from above assumptions we get

$$h'(k_t) = \frac{sB\beta k^{\beta-1}}{1+n} > 0, \quad h''(k_t) = \frac{sB\beta(\beta-1)k^{\beta-2}}{1+n} < 0.$$

The h-map has two fixed point at k = 0 (*trivial and repelling fixed point*) and at $k^s = \left[sB/(1+n)\right]^{1/(1-\beta)}$ (*interior and asymptotically stable*). *The eq. (2.4) is able only to generate monotonic convergence to a fixed point* (See Figure 1.).

If we use the *Leontief* production function , i.e. $f(k_t) = \min\{ak_t, b\} + c$, a,b,c > 0 [Böhm, Kaas, (2000)], that is only piecewise differentiable, from (2.1) we deduce that the dynamical system is described by two affine-linear maps

$$k_{t+1} = h(k_t) = \begin{cases} (ask_t + cs)/(1+n), & \text{if } k_t \le b/a; \\ (bs + cs)/(1+n), & \text{if } k_t > b/a. \end{cases}.$$

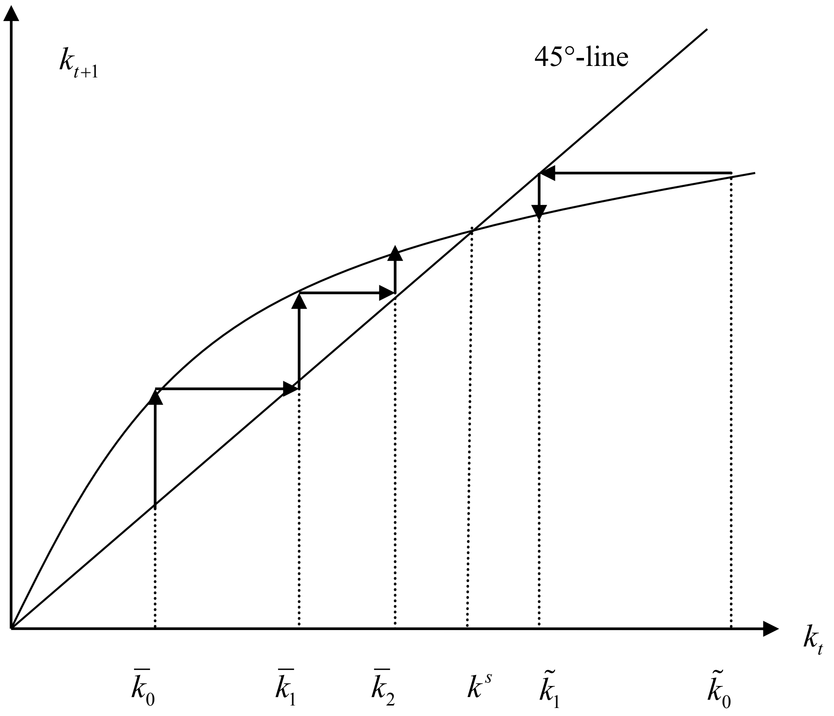


Figure 1: Monotonic Convergence to the fixed point k^s

3. Complex dynamics in the Solow Discrete Time Growth Model

R.H. Day (1982,1983) first has noticed that *complex dynamics can emerge from simple economic structures* as, for example, the neoclassical theory of capital accumulation. Day focuses on the assumptions of the standard Solow growth model and argues that *the kind of nonlinearity of the* $h(k_t)$ *map and the lag present in* (2.1) *are not sufficient to lead to chaos.* Day rewrites (2.1) in a more general form:

$$k_{t+1} = \frac{1}{1+n} s(k_t) f(k_t), \qquad (3.1)$$

where $s(k_t)$ is the saving propensity. Then, Day makes changes in (3.1) deriving two alternative models able to generate chaos in the Li-Yorke (1975) sense. In the first model he keeps $s(k_t)$ as an exogenous constant and modifies the production function $f(k_t)$ into a unimodal map, i.e. a concave and one humped shaped map. Instead in the second model he modifies $s(k_t)$ into a unimodal map and he keeps $f(k_t)$ as an neoclassical production function like the Cobb-Douglas, obtaining a *robust* result [Boldrin and Woodford, (1990)].

In particularly, in the former case he sets $s(k_t) \equiv s$ and defines

$$f(k_t) = \begin{cases} B\, k_t^{\beta} (m - k_t)^{\gamma}, & \text{if } k_t < m; \\ 0, & \text{otherwise} \end{cases}$$

where m is a positive constant, $0 < \beta < 1$, $0 < \gamma < 1$ and $B > 0$.

In the latter case he sets $f(k_t) = Bk_t^{\beta}$ $(B \geq 2, 0 < \beta < 1)$ and he replaces the constant s with the saving function $s(k_t) = a\,(1 - \frac{b}{r})\frac{k_t}{y_t}$, where $r = f'(k_t) = \beta \frac{y_t}{k_t}$, $a > 0, b > 0$.

Thus from the equation (3.1) we deduce respectively the expressions

$$k_{t+1} = \begin{cases} \frac{1}{1+n} sBk_t^{\beta}(m-k_t)^{\gamma}, & \textit{if } k_t \leq m; \\ 0, & \text{otherwise,} \end{cases} \tag{3.2}$$

and

$$k_{t+1} = \frac{a}{1+n}\, k_t \left[1 - \left(\frac{b}{\beta B}\right) k_t^{1-\beta}\right] \tag{3.3}$$

It is very simple to solve the equation (3.2) when $m = \gamma = \beta = 1$. As a matter of fact we can rewrite it like this

$$k_{t+1} = \frac{1}{1+n} s\,B\,k_t(1-k_t) \tag{3.4}$$

If we set $\mu = \frac{sB}{1+n}$ then the (3.4) becomes the well-known logistic equation

$$k_{t+1} = \mu\, k_t(1-k_t).$$

To obtain chaos as Li-Yorke (1975) occurs that in the interval J, in which the continuous map $h(k_t)$ is defined, exists a point k_c such that $h^3(k_c) \leq k_c < h(k_c) < h^2(k_c)$, where $h^2(k_c) \equiv h(h(k_c))$ and $h^3(k_c) \equiv h(h^2(k_c))$. We restrict our study to eq. (3.2). The steps followed by Day to obtain k_c are: (1) he derives the point k^* that maximizes $h(k_t)$ and calls k^m the maximum of $h(k_t)$; (2) he solves the equation $h(k_t) = k^*$ and indicate with k_c the smallest root; (3) he assumes that $h(k^m) = 0$, $k^m < m$ and $k^* < k^m$; (4) he names k^s the steady-state of $h(k_t)$; (5) he observes that, fixing the parameters β, γ and m, *the graph of* $h(k_t)$ *stretches upwards as B is increased* and at same time the position of k^c does not change because in the expression of k^c the parameter B does not appear while the maximum $h(k^c)$ depends linearly on B (See Figure 2 and Figure 3).

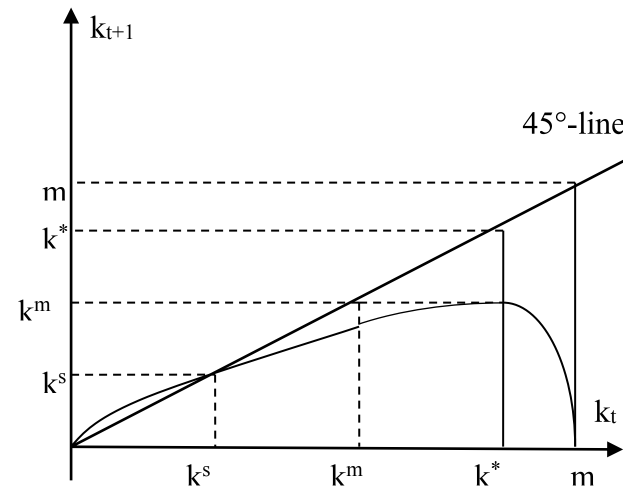


Figure 2: Monotonic Growth or Contraction Chaos

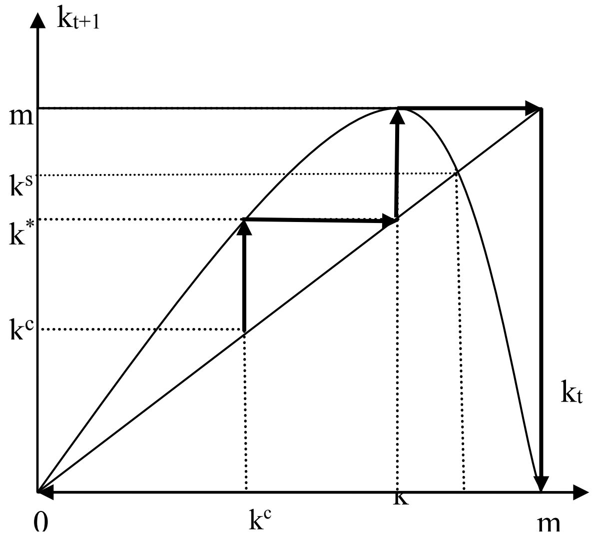


Figure 3: Sufficient Conditions for Li-Yc

4. A Two Class Growth Model: Böhm and Kaas (2000)

4.1 Introduction

The main results of Böhm and Kaas (2000) work are two. The first consists in proving that, slightly changing the standard assumptions of the neoclassical production function (introducing conditions which are slightly weaker than the *Inada conditions*), the dynamics is similar to the one generated by the Solow model. As a matter of fact they define a differentiable, monotonically strictly increasing and strictly concave capital accumulation map that admits a fixed point but not cycles (See below **Proposition 4.1** , **Proposition 4.2**, **Proposition 4.3**).

The second involves the introduction of a Leontief production function, which does not satisfy the weak Inada conditions, in order to construct a piecewise differentiable capital accumulation map. This map can admit zero, one or two steady-states (See **Proposition 4.2.1**, **Proposition 4.2.2**, **Figure 4**. **Figure 5**). Moreover Böhm and Kaas (2000) are also able to derive Li-Yorke chaos [Böhm and Kaas, (2000)].

In particular, in the model of Böhm and Kaas (2000) there are two types of agents (*two class model*), called workers and shareholders, and only one good (or commodity) is produced which is consumed or invested (*one sector model*). Like Kaldor (1956,1957) and Pasinetti (1962), the workers and shareholders have constant savings propensities, denoted respectively with s_w and s_r $(0 \le s \le 1$ and $0 \le s \le 1)$. The output is produced with two factors: labor and capital. We consider that the capital depreciates at a rate $0 < \delta \le 1$ and the labor grows at rate $n \ge 0$. We write the production function $f: \Re \to \Re$ in intensive form (i.e. it is maps capital per worker k into output per worker y), and suppose that f satisfies the following conditions :

- f is C^2;
- $f(\lambda k) = \lambda f(k)$ (*constant returns to capital*);
- f is monotonically increasing and strictly concave (i.e. $f'(k)>0$ and $f''(k)<0$ for all k > 0);
- $\lim_{k\to\infty} f(k) = \infty$;
- $\lim_{k\to 0} \frac{f(k)}{k} = \infty$ and (b) $\lim_{k\to\infty} \frac{f(k)}{k} = 0$ (*weak Inada conditions (WIC)*)

If we assume that the market is competitive then the wage rate w(k) is coincident with the marginal product of labor, i.e. $w(k) = f(k) - k f'(k)$, and the interest rate (or investment rate) r is equal to the marginal product of capital, i.e. $r = f'(k)$. We suppose that f(0) generally is not equal to 0. We observe that the total capital income per worker is $kf'(k)$. Moreover from WIC we deduce that:

- $w(k) \ge 0$;
- $w'(k) = - k f''(k) > 0$ (w(k) is strictly monotonically increasing);
- $0 \le kf'(k) \le f(k) - f(0)$;
- $\lim_{k\to 0} kf'(k) = 0$.

Similarly to the Solow model we obtain that the time-one map of capital accumulation is

$$k_{t+1} = G(k_t) = \frac{1}{1+n}((1-\delta)\, k_t + s_w w(k_t) + s_r k_t f'(k_t)) \quad (4.1).$$

Proposition 4.1.1 Given $n \ge 0$ and $0 \le \delta \le 1$, let f(k) be a production function which satisfies the WIC. If the workers do not save less than shareholders (i.e. $s_w \ge s_r$) or $e_{f'(k)} \ge -1$ then G is monotonically increasing in k.

The following proposition investigates *the existence and the uniqueness of steady states*.

Proposition 4.1.2 Consider n and δ fixed and let f(k) be a production function which satisfies the WIC. The following conditions hold:

- k = 0 if and only if $s_w = 0$ or f(0) = 0.

- There exists al least one positive steady state if $(s_r > 0 \text{ and } \lim_{k\to 0} f'(k) = 0)$ or if $(s_w > 0 \text{ and } f'(0) < \infty)$.
- There exists at most one positive steady state if $(s_r \geq s_w)$.

Proposition 4.1.3 k^* is a steady state of Pasinetti-Kaldor iff, for given n and δ the pairs (s_r, s_w) of savings rates describe the line $s_r + \frac{1-e_f(k^*)}{e_f(k^*)} s_w = 1$ in the (s_r, s_w)-diagram, where $e_{f(k)} = \frac{kf'(k)}{f(k)}$

- has negative slope;
- goes across the point $(s_r, s_w) = (1,0)$;
- is below or above the 45°- line $s_w = s_r$ depending on $e_{f(k^*)}$ is less or greater than $1/2$.

The (s_r, s_w) - plane is coincident with the square $[0,1]^2$.

4.2 The dynamics with fixed proportions

We consider the Leontief technology $f_L(k) = \min\{ak, b\} + c, \quad a, b, c > 0$.

Let $k^* = b/a$ be. We have

$$f_L(k) = \begin{cases} ak+c, & \text{if } k \leq k^*, \\ b+c, & \text{if } k > k^* \end{cases} \quad \text{and} \quad f'_L(k) = \begin{cases} a, & \text{if } k \leq k^*; \\ 0, & \text{if } k > k^*. \end{cases}$$

The map G becomes

$$G_L(k) = \begin{cases} G_1(k) = \frac{1}{1+n}((1 - \delta + s_r a)\, k + s_w c), & \text{if } k \leq k^*, \\ G_2(k) = \frac{1}{1+n}((1-\delta)k + (b+c)\, s_w), & \text{if } k > k^*. \end{cases}$$

We may say that:

- G_1 and G_2 are affine-linear maps strictly monotonically increasing;
- $G'_1 = \frac{1}{1+n}(1 - \delta + s_r a) > G'_2 = \frac{1}{1+n}(1 - \delta)$;
- $G'_2 < 1$: the map G'_2 has always a fixed point k_2 if we define $G_2(k)$ for all $k \geq 0$;
- G_1 has the fixed point k_1 if and only if $G'_1 < 1$, that is $n + \delta - s_r a > 0$;
- $G_1(0) = \frac{1}{1+n} s_w c < G_2(0) = \frac{1}{1+n}(b+c)\, s_w$ if we define $G_2(k)$ for all $k \geq 0$.

Let k_1 be the fixed point for G_1. Then k_1 *is a fixed point also for G if and only if* $k_1 < k^*$. Analogously, found the fixed point k_2 for G_2, we have that k_2 *is a fixed point also for* G *if and only if* $k^* < k_2$.

Proposition 4.2.1 Let $G'_1 < 1$ be. We obtain that:

(i) the fixed point k_1 for G_1 is equal to $\frac{cs_w}{n+\delta-as_r}$;

(ii) k_1 is a fixed point also for G if and only if $bs_r + cs_w < (n+\delta)\frac{b}{a}$;

(iii) $G_1(k^*) < k^*$ if and only if $bs_r + cs_w < (n+\delta)\dfrac{b}{a}$.

Proposition 4.2.2 We get

(i) the fixed point of G_2 is $k_2 = \dfrac{(b+c)s_w}{n+\delta}$;

(ii) k_2 is the fixed point also for G if and only if $s_w > \dfrac{(n+\delta)b}{(b+c)a}$;

(iii) $G_2(k_*) > k^*$ if and only if $s_w > \dfrac{(n+\delta)b}{(b+c)a}$.

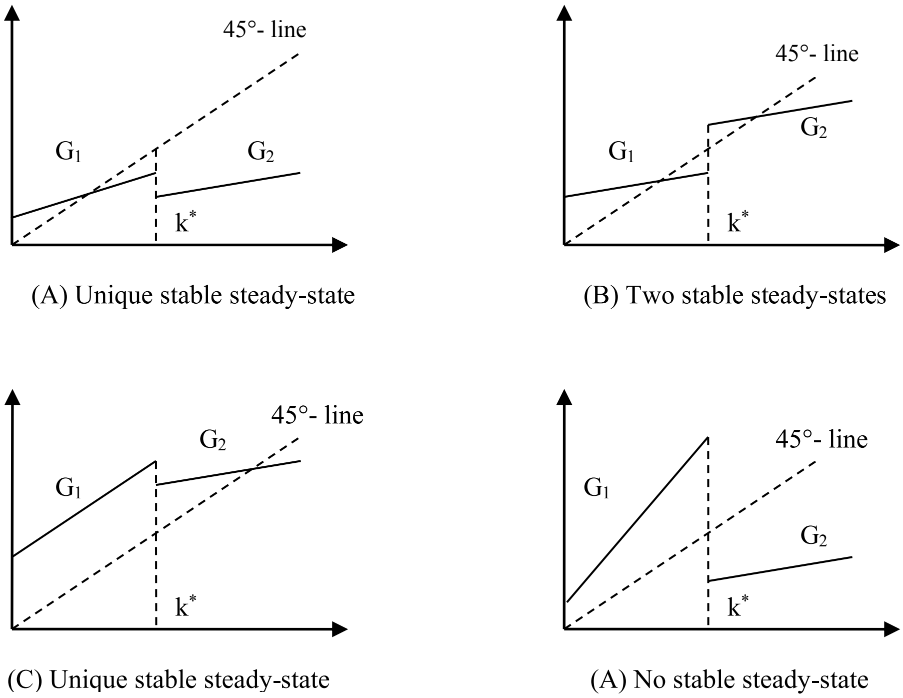


Figure 4 Types of time-one map with Leontief technology

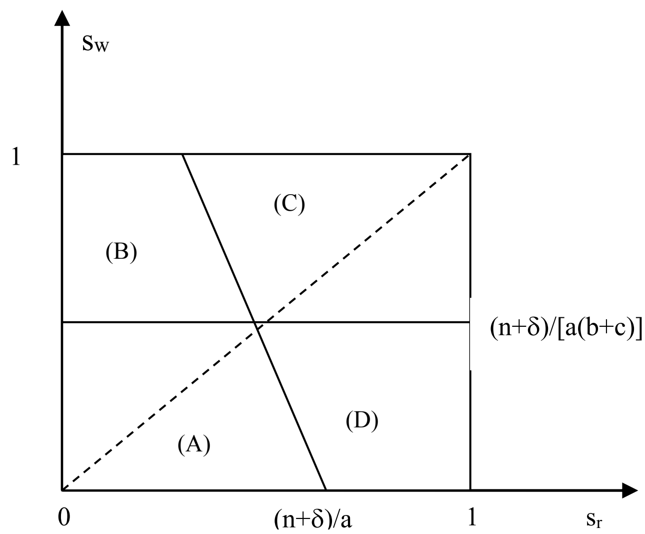


Figure 5 Stability regions for the Leontief technology

5. Complex Dynamics in a Pasinetti-Solow Model of Growth and Distribution: a Model of P.Commendatore

5. 1 Introduction

Similarly to the paper of Böhm and Kaas (2000), the model of Commendatore (2005)

- is a two-class model, that is two distinct group of economic agents (workers and capitalists) exist, with constant propensities to save [Kaldor, (1956)];
- labor and capital markets are perfectly competitive;
- the income sources of workers are wages and profits and the income of capitalists is only profits [Pasinetti, (1962)];
- the time is discrete;
- there is a single good in the economy (one sector model).

Commendatore's model differs from the model of Böhm and Kaas in some assumptions:

- following Chiang (1973), workers not save in same proportions out of labor and income of capital;
- the production function is not with fixed proportions (Leontief technology) but it is a CES production function;
- likewise Samuelson-Modigliani (1966) that, following Pasinetti (1962), extend the Solow growth model (1956) to two-dimensions, the map that describes the accumulation of capital in discrete time is two-dimensional because it considers not only the different saving behaviour of two-classes but also their respective wealth (capital) accumulation.

5.2 The model: the economy, short-run equilibrium, steady growth equilibrium

Let $f(k)=\left[\alpha+(1-\alpha)k^{\rho}\right]^{1/\rho}$ be the CES production function in intensive form, where k is the capital/labor ratio, $0<\alpha<1$ is the distribution coefficient, $-\infty<\rho<1$ ($\rho\neq 0$), $\eta=1/(1-\rho)$ is the constant elasticity of substitution. We consider $f(k)>0$. Therefore $f(k)=\left[\alpha+(1-\alpha)k^{\rho}\right]^{1/\rho}=\left[\alpha k^{-\rho}+(1-\alpha)\right]^{1/\rho}k$. The terms k_w and k_c denote, respectively, workers' and capitalists' capital per worker, where $0\leq k_w\leq k$, $0\leq k_c\leq k$, $k=k_w+k_c$. The workers' saving out of wages are represented by $s_{ww}(f(k)-kf'(k))$ and the workers' saving out of capital revenues consist in $s_{wP}f'(k)k_w$, where $0\leq s_{ww}\leq 1$, $0\leq s_{wP}\leq 1$. Instead the capitalists' savings are $s_cf'(k)k_c$, where $0\leq s_c\leq 1$. We assume $s_c>\max\{s_{ww},s_{wP}\}$. Thus the aggregate savings correspond to $s(k_c,k_w)=s_{ww}(f(k)-f'(k)k)+s_{wP}f'(k)k_w+s_cf'(k_c)$.

Let n be the constant rate of growth of labor force, the following map $G(k_w,k_c)=\frac{1}{1+n}\left[(1-\delta)k+i\right]$ describes the rule of capital accumulation per worker, where i indicates gross investment per worker and $0<\delta<1$ is the constant rate of capital depreciation. In a short-run equilibrium G becomes

$$G(k_w,k_c)=\frac{1}{1+n}\left[(1-\delta)k+s_{ww}(f(k)-f'(k)k)+s_{wP}f'(k)k_w+s_cf'(k_c)\right], \qquad (5.2.1)$$

from which we deduce the capitalist' process of capital accumulation

$$G_w(k_w,k_c)=\frac{1}{1+n}\left[(1-\delta)k_w+s_{ww}(f(k)-f'(k)k+s_{wP}f'(k)k_w\right] \qquad (5.2.2)$$

and the capitalist's rule of capital accumulation

$$G_c(k_w,k_c)=\frac{1}{1+n}\left[(1-\delta)k_c+s_cf'(k)k_c\right] \qquad (5.2.3)$$

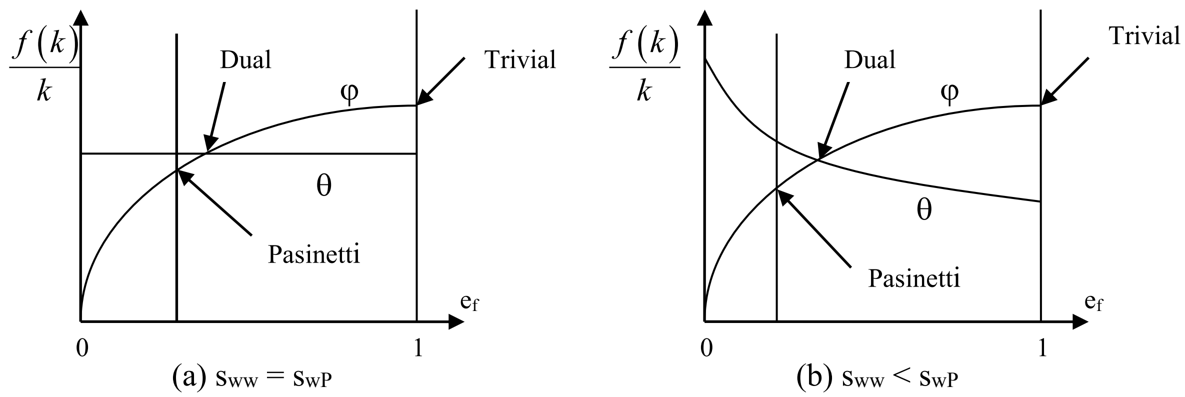
In order to obtain the steady states of G_w and G_c, we imposing

$G_w(k_w,k_c)=k_w$ and $G_c(k_w,k_c)=k_c$.

We get $(n+\delta)k_w = s_{ww}(f(k) - f'(k)k) + s_{wP}f'(k)k_w$, $(n+\delta)k_c = s_c f'(k_c)$.

Commendatore (2005) find three types of equilibria: *Pasinetti -equilibrium* (capitalists own positive share of capital), *dual equilibrium* (only workers own capital) and *trivial equilibrium* (the overall capital is zero) and, developing ingeniously a geometrical method used by Meade [Meade, (1966)], describes geometrically the coexistence of them (See Figure 6). To detect above equilibria he relates $\frac{f(k)}{k}$ to $e_{f(k)} = \frac{f'(k)k}{f(k)}$ and he finds that $\frac{f(k)}{k} = \varphi(e_f)$ for Pasinetti-equilibrium and $\frac{f(k)}{k} = \theta(e_f)$, where $\varphi(x) \equiv \left(\frac{1-\alpha}{x}\right)^{1/\rho}$ $(0 < \alpha < 1)$ and $\theta(x) \equiv \frac{n+\delta}{s_{ww}(1-x) + s_{wP}x}$.

For example, if $\rho < -1$, the curve $\varphi(e_f)$ is monotonically strictly increasing and strictly concave while the curve $\theta(e_f)$ is (a) a horizontal line if $s_{ww} = s_{wP}$, (b) monotonically strictly decreasing and strictly concave if $s_{ww} < s_{wP}$ and (c) monotonically strictly increasing and strictly convex if $s_{ww} > s_{wP}$. Look at the Figure 6: the intersection between $\varphi(e_f)$ and $\theta(e_f)$ represents a Dual-equilibrium, the point $(1, \varphi(1))$ gives a trivial equilibrium and the intersection between the vertical line $e_f(k^P) = (1-\alpha)^{\frac{1}{1-\rho}} \left(\frac{n+\delta}{s_c}\right)^{\frac{\rho}{\rho-1}}$ and $\varphi(e_f)$ identifies the Pasinetti -equilibrium. To derive the remaining cases we observe that the other diagrams of $\varphi(x)$ are showed in Figure 7.



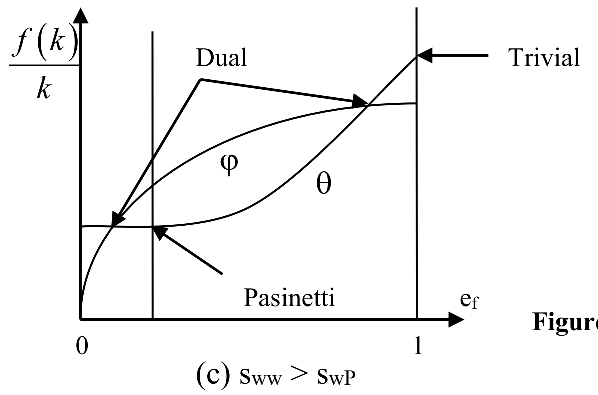


Figure 6: Steady growth equilibria are identified for the cases (a) $s_{ww} = s_{wP}$, (b) $s_{ww} < s_{wP}$, (c) s_{ww}

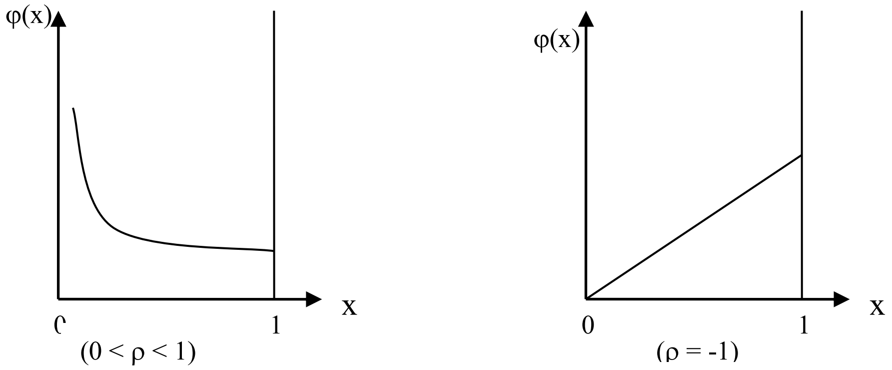


Figure 7 The diagram of $\varphi(x)$ for (0 < ρ < 1) and (ρ = -1).

The author analyses the *local stability* and the *global stability* of the nonlinear dynamical system given by (5.2.2) and (5.2.3). To study the local stability, he starts from Pasinetti equilibria and he finds the trace **T** and the determinant **D** of the Jacobian matrix **J** evaluated at a Pasinetti-equilibrium fixed. Then he applies the *conditions of stability* of dynamical system [Azariadis, C., (1993)] and identifies the stability region (*Triangle Stability*) in TD-plane. Moreover, with the Theory of Local Bifurcations, he studies which bifurcation appears when a given Pasinetti-equilibrium loses a stability. Using the *characteristic equation* he derives the eigenvalues of **J** and analyses the global stability.

6. Conclusions

We observe that Commendatore's model generalizes Böhm and Kaas (2000) model and Solow (1956) model. As a matter of fact

- setting $s_{ww} = s_{wP}$ and $k = k_w = k_c$ in (5.2.1)

$$G(k_w,k_c)=\frac{1}{1+n}\left[(1-\delta)k + s_{ww}(f(k) - f'(k)k) + s_{wP}f'(k)k_w + s_c f'(k_c)\right],$$

we have the (4.1), i.e. from Commendatore's model we deduce Böhm and Kaas (2000) model;

- setting $s_w = s_r$ in (4.1)

$$k_{t+1} = G(k_t) = \frac{1}{1+n}((1-\delta)\, k_t + s_w w(k_t) + s_r\, k_t\, f'(k_t)),$$

we obtain the (2.2), i.e. from Böhm and Kaas (2000) model we deduce the Solow (1956) model.

References

[1] Alligood, K.T., Sauer, T.D., Yorke, J.A, (1996), *Chaos An Introduction to Dynamical Systems*, Springer.

[2] Azariadis, C., (1993), *Intertemporal Macroeconomics*, Blackwell.

[3] Böhm, V., Kaas, L., (2000), *Differential savings, factor shares, and endogenous growth cycles*, in: *Journal of Economic Dynamics & Control*, 24, pp. 965 – 980.

[4] Böhm, V., Pampel, T., Wenzelburger, J., (2007), *On the Stability of Balanced Growth*, Discussion Paper No. 548, Bielefeld University.

[5] Benhabib, J., (1991), *Cycles and Chaos in Economic Equilibrium*, Princeton University Press.

[6] Boldrin, M. and Woodford, M., (1990), *Equilibrium Models Displaying Endogenous Fluctuations and Chaos: A Survey*, in: *Journal of Monetary Economics*, 25, pp. 1989 – 1990.

[7] Chiang, A.C., (1973), *A simple generalization of the Kaldor-Pasinetti theory of profit rate and income distribution*, in: *Economica*, New Series, Vol.40, No. 4, pp. 311 – 313.

[8] Commendatore, P., (2005), *Complex Dynamics in a Pasinetti-Solow model of growth and distribution*, forthcoming in: *Macroeconomics 0511006*, *EconWPA*; forthcoming in *Economics & Complexity*.

[9] Day, R.H., (1982), *Irregular growth cycles*, in: *American Economic Review*, 72, pp. 406 – 414.

[10] Day, R.H., (1982), *The emergence of Chaos from Classical Economic Growth*, in: *The Quaterly Journal of Economics*, Vol. 98, No. 2, pp. 201 – 213.

[11] Drandakis, E.M., (1963), *Factor Substitution in the Two-Sector Growth Model*, in: *The Review of Economic Studies*, Vol. 30, N0. 3, pp. 217 – 228.

[12] Hahn, F.H., (1955), *On Two-Sector Growth Models*, in: *Review of Economic Studies*, 32, 4, pp. 339 – 346.

[13] Kaldor, N., (1956), *Alternative theories of distribution*, in: *Review of Economic Studies*, 23, pp. 83 – 100.

[14] Kaldor, N., (1957), *A model of Economic Growth*, in: *The Economic Journal,* 67, pp. 591 – 624.

[15] Kurz, M., (1963), *A Two-Sector Extension of Swan's Model of Economic Growth: the case of no Technical Change*, in: *International Economic Review*, Vol. 4, No. 1, pp. 69 – 79.

[16] Li,T.Y.,Yorke, J.A., (1975), *Period three implies chaos*, in: *American Mathematical Monthly*, Vol. No. 82, pp. 985 – 992.

[17] Meade, J.E, (1961), *A Neoclassical Theory of Economic Growth*, New York: Oxford University Press.

[18] Meade, J.E., (1966), *The outcome of the Pasinetti-process a note*, Economic Journal, pp.161–164.

[19] Oniki, H., and Uzawa, H., (1965), *Patterns of Trade and Investment in a Dynamic Model of International Trade*, in: *The Review of Economic Studies*, Vol.34, No. 2, pp. 227 – 238.

[20] Pasinetti, L., (1962), *Rate of Profit and Income Distribution in relation to Rate of Economic Growth*, in: *The Review of Economic Studies*, Vol. 29, No. 4, pp. 267 – 279.

[21] Ramsey, F., (1928), *A Mathematical Theory of Saving* , *Economic Journal*, 38, 152, pp.543– 559.

[21] Samuelson, P.A., Modigliani, F., (1966), *The Pasinetti Paradox in neoclassical and more general models*, 33, pp. 269 – 301.

[22] Shinkai, Y., (1960), *On Equilibrium Growth of Capital and Labor*, in: *International Economic Review*, Vol. 1, No. 2, pp. 107 – 111.

[23] Srivasan, T.N., (1964), *Optimal Savings in a Twp-Sector Model of Growth*, in: *Econometrica*, Vol. 32, No. 3, pp. 358 – 373.

[24] Solow, R.M., (1956), *A contribution to the theory of economic growth*, in: *Quaterly Journal of Economics,* 70, pp. 65 – 94.

[25] Stiglitz, J.E., (1967), *A Two-Sector Two-Class Model of Economic Growth*, in: *The Review of Economic Studies,* Vol. 34, N0. 2, pp. 227 – 238.

[26] Takajama, A., (1963), *On a Two-Sector Economic Growth Model with a Technological Progress. A Comparative Statics Analysis*, in: *The Review of Economic Studies*, Vol. 30, 95 – 104.

[27] Takajama, A., (1965), *On a Two-Sector Economic Growth Model with a Technological Progress. A Comparative Statics Analysis*, in: The Review of Economic Studies, Vol.32, No. 3, pp. 251-262.

[28] Uzawa, H., (1961), *On a Two-Sector Model of Economic Growth*, in: *The Review of Economic Studies,* Vol. 29, No.1, pp. 40 – 47.

[29] Uzawa, H., (1963), *On a Two-Sector Model of Economic Growth II*, in: *The Review of Economic Studies*, Vol. 30, No.2, pp. 105 – 118.

AN EXPLORATORY STUDY OF INDIAN FOREIGN TRADE

Rajesh K **PILLANIA**
Management Development Institute, **INDIA**
r_pillania@yahoo.com

Abstract

Indian economy and foreign trade are on a growth trajectory. Indian exports have come a long way in value terms from the time of gaining independence in 1947. The total value of India's merchandise exports increased from US $ 1.3 billion in 1950-51 to US $ 63.8 billion in 2003-04 – a compound rate of 7.6 per cent. Trade growth has picked up post liberalization of 1991. The composition of trade is now dominated by manufactured goods and services. India services exports share in global exports is more than double of that of Indian manufacturing exports. East Asian countries, particularly China have become a major trading block. There is huge untapped potential for Indian foreign trade in years to come.

Keywords: india, foreign trade, liberalization, trade composition, trade direction.

JEL Classification: F14, F41, F43

1. Introduction

Indian exports have come a long way from the time of independence in terms of value. The total value of India's merchandise exports increased from US $ 1.3 billion in 1950-51 to US $ 63.8 billion in 2003-04 – a compound rate of 7.6 per cent [Malik, (2005)].Indian economy and foreign trade has shown progress post liberalization. In contrast to the pre-reform period (1950-90), the actual growth of exports in the post-reform period has been above the potential offered by the growth of world demand. The gap between the actual and potential is mainly explained by an improvement in the overall competitiveness of India's exports [Virmani, (2003), Veeramani, (2007)].

Over the last few years, the growth rates has picked up. The current account has followed an inverted "U" shaped pattern during the period from 2001-02 to 2006-07, rising to a surplus of over 2 per cent of GDP in 2003-04. Thereafter it has returned close to its post-1990s reform average, with a current account deficit of 1.2 per cent in 2005-06 and 1.1 per cent of GDP in 2006-07. Capital inflows, as a proportion of GDP, have been on a clear uptrend during the six years (2001-02 to 2006-07) of this decade. They reached a high of 5.1 per cent of GDP in 2006-07 after a somewhat modest growth rate of 3.1 per cent in 2005-06. The net result of these two trends has been a gradual rise in reserve increase to reach 4 per cent of GDP in 2006-07.

With capital inflows exceeding financing requirements, foreign exchange reserve increase was of the order of US$ 15.1 billion in 2005-06 and US$ 36.6 billion in 2006-07.As a proportion of GDP, external debt was 17.2 per cent and 17.9 per cent in 2005-06 and 2006-07 respectively [Ministry of Finance, (2008)]. This research paper studies Indian foreign trade since 1949. It consists of three sections including this introduction part. The second section studies various aspects of Indian foreign trade and the last section is conclusion part.

2. Indian Foreign Trade over the Years 1949-2006

Indian foreign trade has grown in absolute numbers as compared to 1950-51, but its share in world trade has gown down from around 2.5 percent to 0.67 percent in 1991 and increased to more than one percent in 2007. For the purpose of study, foreign trade can be divided into three periods namely 1950-1970, 1971-1991 and post 1991.

During the first phase, 1950-1970, exports have grown at a very slow rate. During 1950s the exports growth rate was 3.6 percent in dollar terms and 3.5 percent in 1960s as shown in Table 1.Due to rising imports and stagnant exports, policy of import substitution was started in 1960s to cut down on imports. Five primary commodities constituted a major portion of Indian exports and the prevailing belief was that the country had nothing much to export. Government had adopted a policy of export pessimism and import substitution during this period. Exports were largely neglected during the first and the second five-year plans, which was justified on the ground that demand for Indian exports was inelastic. Whilst the world merchandise export was growing at 6.3 per cent per annum during the 1950s, exports from India stagnated. As the world merchandise exports expanded relatively faster during the 1960s at 8.8 per cent per annum, the growth rate of India's exports improved somewhat to 3.6 per cent per annum. Clearly, the country failed to make the best use of the trade possibilities available during the 1950s and 1960s [Singh, (1964); Bhagwati and Srinivasan, (1975); Nayyar, (1976); Veeramani, (2007)]. Several studies have argued that the import substitution policies had created a bias against exports in India. In spite of the various export promotion schemes adopted in the 1970s and 1980s, profitability in the heavily protected domestic market remained significantly higher than that in the export market [Kathuria (1996), Veeramani, (2007)].

Table1. Performance of Trade Sector (Export and Import) in India

Performance of Trade Sector (Export and Import) in India (1950-1951 to 1999-2000)									
	Growth Rate								
Annual Average	In Rupee Terms		In US Dollar terms		In Real terms (Volume)#		As per cent to GDP		
	Exports	Imports	Exports	Imports	Exports	Imports	Exports	Imports	(Exports+ Imports)@
1950-51 to 1959-60	3.6	6.1	3.6	6.1	3.0	4.1	5.1	6.7	11.8
1960-61 to 1969-70	8.9	6.3	3.5	0.8	2.7	1.4	3.4	5.5	8.9
1970-71 to 1979-80	16.8	20.7	15.8	20.1	7.3	6.7	4.5	5.3	9.8
1980-81 to 1989-90	16.4	14.9	8.0	7.2	4.2	7.3	4.6	7.2	11.8
1990-91 to 1999-2000	19.5	20.1	8.6	9.6	11.0	12.2	7.8	9.3	17.1
1992-93 to 1995-96	24.7	26.8	15.7	17.5	17.5	21.3	8.1	9.0	17.1

Note #: Refers to calendar year. Volume obtained by dividing value of exports (f.o.b) and imports (c.i.f) with their respective unit prices. @: Figures may not add-up due to rounding off.

Source: Report on Currency and Finance (2002-03), Reserve Bank of India; IndiaStat Database (2008).

During the period of 1970-1991 exports performance improved. Government had taken initiatives in late 1960s like establishing Indian Institute of Foreign Trade and others for promoting foreign trade. The world economy was also growing fast in 1970s. The export growth rate was 15.8 percent in 1970s before slowing down to 8 percent in 1980s.During 1970s, imports growth rate also picked up and infact was higher than growth rate of exports. The contribution of foreign trade to GDP again reached to 11.8 per cent, the same level as on 1950-51. The export boom of the 1970s, however, could not be maintained during the first half of the 1980s. As the growth rate of world exports turned negative in the aftermath of the second oil price hike, India's exports decelerated sharply. During the second half of the 1980s, however, the world economy recovered and India's exports grew at a healthy pace (17.8 per cent). There was a genuine improvement in the export competitiveness of India during this period due to a major depreciation of the REER and increased export subsidies. This period also witnessed some doses of industrial deregulation and liberalization of capital goods imports [Joshi and Little (1994); Veeramani, (2007)].

In the post liberalization period i.e. post 1991, export and import growth has picked up and the contribution of foreign trade to GDP has increased to 17.1 percent by 2000.However during the period import growth rates has been higher than exports growth rates. Many pro export policies were started after liberalization. Export promotion schemes prevalent during the post 1991 period include: export promotion capital goods (EPCG), duty entitlement passbook (DEPB), duty free replenishment certificate (DFRC), advance licences, special import licence (SIL), exemption from income tax, sector/market-specific schemes [e g, market access initiative (MAI), towns of export excellence, agri export zones (AEZ), Focus Africa, and Focus Latin American Countries], and schemes for statusholders, export oriented units (EOUs), units in special economic zones (SEZs), electronic hardware technology parks (EHTPs), software technology parks (STPs) and biotechnology parks (BTPs). A few more schemes (such as, target plus, served from India) have been added under the Foreign Trade Policy 2004 [RBI (2004), Malik, (2005)].

Table 2. Export/Import and Trade Balance of India

Export/Import and Trade Balance of India (1949-1950 to 1990-91) (US $ Million)			
Year	Export (including Re-export)	Import	Trade Balance
1949-50	1016	1292	-276
1950-51	1269	1273	-4
1951-52	1490	1852	-362
1952-53	1212	1472	-260
1953-54	1114	1279	-166
1954-55	1233	1456	-223
1955-56	1275	1620	-345
1956-57	1259	1750	-491
1957-58	1171	2160	-989
1958-59	1219	1901	-682
1959-60	1343	2016	-674
1960-61	1346	2353	-107
1961-62	1381	2281	-900
1962-63	1437	2372	-935
1963-64	1659	2558	-899
1964-65	1701	2813	-1111
1965-66	1693	2944	-1251
1966-67	1628	2923	-1295
1967-68	1586	2656	-1071
1968-69	1788	2513	-726
1969-70	1866	2089	-223
1970-71	2031	2162	-131
1971-72	2153	2443	-290
1972-73	2550	2415	134
1973-74	3209	3759	-549
1974-75	4174	5666	-1492
1975-76	4665	6084	-1420
1976-77	5753	5677	76
1977-78	6316	7031	-715
1978-79	6978	8300	-1322
1979-80	7947	11321	-3374
1980-81	8486	15869	-7383
1981-82	8704	15174	-6470
1982-83	9107	14787	-5679
1983-84	9449	15311	-5861
1984-85	9878	14412	-4534
1985-86	8904	16067	-7162
1986-87	9745	15727	-5982
1987-88	12089	17156	-5067
1988-89	13970	19497	-5526
1989-90	16612	21219	-4607
1990-91	18145	24073	-5927

Source: Handbook of Industrial Policy and Statistics, 2001, Ministry of Commerce & Industry and Department of Economic Affairs, Govt. of India & Monthly Newsletter, Vol. XL, No. 10, October 25, 2004, Indian Investment Centre; IndiaStat Database(2008).

The trade balance has always been negative as shown in Table 2 and Table 3 except two years 1972-73, 1976-77.The trade deficit has been increasing in recent years.

Table 3. Export/Import and Trade Balance of India

Trends of Foreign Trade (US $) in India (1990-1991 to 2006-2007) (US $ Million)					
Year	Export	Growth Rate	Import	Growth Rate	Trade Deficit
1990-91	18145	9.2	24072	13.4	-5927
1991-92	17865	-1.5	19411	-19.4	-1546
1992-93	18437	3.7	21882	12.7	-3345
1993-94	22237	20	23306	6.5	-1069
1994-95	26330	18.4	28654	22.9	-2324
1995-96	31797	20.8	36678	28	-4881
1996-97	33470	5.3	39132	6.7	-5662
1997-98	35006	4.6	41484	6	-6478
1998-99	33219	-5.1	42389	2.2	-9170
1999-2000	36822	10.8	49671	17.2	-12848
2000-01	44560	21	50536	1.7	-5976
2001-02	43827	-1.65	51413	1.74	-7587
2002-03	52719	20.29	61412	19.45	-8693
2003-04	63843	21.10	78150	27.25	-14307
2004-05	83536	-	111518	-	-27982
2005-06	103091	-	149166	-	-46075
2006-07 (P)	89489	-	131212	-	-41723

Source: Ministry of Commerce & Industry, Govt. of India; IndiaStat Database (2008).

During post liberalization era, exports have done well particularly from 1992-93 to 1996-97; and from 2002-2003 to 2007-2008. As a proportion of GDP, on balance of payments (BoP) basis, exports rose from a level of 5.8 per cent in 1990- 91 to reach a level of 14.0 per cent of GDP in 2006-07. The average annual growth rate in the last five years has been placed at a high of 23.5 per cent. However, imports have grown even faster in the last five years at an annual average of 28.2 per cent. As a proportion of GDP, on BoP basis, imports in 2006-07 were placed at 20.9 per cent of GDP. Thus, trade deficit widened to 6.9 per cent of GDP in 2006-07. The higher trade deficit could be attributed to a rise in petroleum, oil and lubricants (POL) as well as non-POL components in imports. Continued uptrend in prices in the international markets and rise in the price of gold were the major contributors to this process [Finance Ministry, (2008)]. The trade account is supported by the rising services exports. India's services exports, at $81.3 billion (Rs3.2 trillion) in 2006-07, are fast catching up with the country's merchandise exports of $127.1 billion. The services export growth rate in 2006-07 was 32.5% compared to 21% in merchandise export [Singh,(2007)]. IBEF has estimated that if the average annual growth rates of the last three years – 56.3 per cent for service exports and 21.8 per cent for manufacturing exports – were projected into the future, by the beginning of 2007, services could topple merchandise goods at the pole position in exports [IBEF, (2005)]. Indian share of global services export is more than double of merchandise exports and India is one of the few countries which have increased their share of services exports in recent years.

Private transfers receipts (mainly remittances) shot up, year-on-year, by 49.2 per cent as against 19.2 per cent in the corresponding period of the previous year. Investment income (net) grew by 60.0 per cent in 2007-08 (April-September) reflecting the burgeoning foreign exchange reserves. Net invisible surplus grew by 35.2 per cent to reach US$ 31.7 billion in 2007-08 (April-September), equivalent of 6.1 per cent of GDP. Thus, higher invisible surplus was able to moderate somewhat the rising deficits on trade account and current account deficit was placed at US$ 10.7 billion in 2007-08 (April-September), equivalent of 2.0 per cent of GDP [Finance Ministry, (2008)].

Table 4 shows the Terms of Trade for Indian economy from 1969-1970 to 2005-2006.The terms of trade has been fluctuating over the years.

Table 4. Index Number of Foreign Trade

Index Number of Foreign Trade (Base : 1978-79=100) (1969-1970 to 2005-2006)							
Year	Unit Value Index		Volume Index		Terms of Trade		
	Exports	Imports	Exports	Imports	Gross	Net	Income
1969-70	44.0	35.2	55.7	64.9	116.5	125.0	69.6
1970-71*	45.0	35.3	59.0	67.2	113.9	127.4	75.2
1971-72*	46.0	32.8	59.2	80.6	136.1	140.2	83.0
1972-73*	51.2	34.2	66.5	76.7	115.3	149.7	99.6
1973-74*	62.2	48.9	69.5	87.2	125.4	127.2	88.4
1974-75*	78.0	84.5	73.7	77.2	104.7	92.3	68.4
1975-76*	83.9	99.1	81.7	76.0	93.0	84.7	69.2
1976-77*	89.4	96.3	96.8	76.1	78.6	92.9	89.9
1977-78*	100.3	88.0	93.2	100.0	107.3	114.0	106.2
1978-79	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1979-80	105.4	114.1	106.2	116.4	109.7	92.4	98.1
1980-81	108.5	134.2	108.1	137.9	127.6	80.8	87.3
1981-82	124.1	133.1	110.1	150.6	136.8	93.2	102.6
1982-83	132.0	136.3	116.7	154.6	132.4	96.8	113.0
1983-84	151.0	125.8	113.0	185.4	164.1	120.0	135.6
1984-85	169.8	161.7	120.8	156.1	129.2	105.0	126.8
1985-86	170.8	158.8	111.3	182.3	163.8	107.6	119.8
1986-87	179.4	139.4	121.3	212.3	175.0	128.6	156.0
1987-88	195.4	160.0	140.0	204.8	146.3	122.1	170.9
1988-89	232.2	185.5	152.1	224.2	147.4	125.2	190.4
1989-90	276.6	228.4	174.9	227.8	130.2	121.1	211.8
1990-91	292.5	267.7	194.1	237.7	122.5	109.3	212.2
1991-92	369.5	309.1	208.6	228.0	109.3	119.5	249.3
1992-93	421.5	331.0	222.9	282.0	126.5	127.3	283.8
1993-94	474.1	327.2	257.5	329.1	127.8	144.9	373.1
1994-95	494.6	324.6	292.7	408.3	139.5	152.4	446.0
1995-96	484.2	351.0	384.3	514.8	134.0	137.9	530.0
1996-97	504.7	399.8	411.8	511.8	124.3	126.2	519.7
1997-98	589.4	404.2	386.0	562.1	145.6	145.8	562.8
1998-99	611.7	407.8	399.2	644.2	161.4	150.0	598.8
1999-00	604.5	450.5	461.0	704.8	152.9	134.2	618.7
2000-01	624.3	487.5	571.4	697.7	122.1	128.1	732.0
2001-02	618.0	492.9	592.7	732.8	123.6	125.4	743.2
2002-03	619.6	545.6	721.6	802.4	111.1	113.6	819.7
2003-04	672	545	765	970	-	-	-
2004-05	732	663	899	1113	-	-	-
2005-06	798	592	1005	1649	-	-	-

Note: 1. Gross Terms of Trade implies Volume Index of Imports expressed as a percentage of Volume Index of Exports.

2. Net Terms of Trade implies unit value Index of Exports expressed as a percentage of unit value index of Imports.

3. Income Terms of Trade implies the product of Net Terms of Trade and Volume Index of Exports as a percentage.

* : Converted from the original base with the help of linking factors.

Source: Ministry of Commerce & Industry, Govt. of India. & Department of Economics and Statistics Govt. of Tamil Nadu, IndiaStat Database (2008).

Estimated Foreign Trade by 2020

Trend analysis was used to project the export and import growth and trade deficit/ surplus up to 2020. For this, time series data, ranging from 1950-2007 on export, import and trade deficit/surplus was used. The analysis shows that export trade will grow up to US$ 61 billion. On the other hand import will continue to grow and cross US $ 79 billion. The trade deficit will keep increasing and reach USD 18 billion by 2020. Table 5 shows estimated trade till 2020.

Table 5. Estimated Exports, Imports and Trade Deficit by 2020

Year	Estimated Export Y1^	Estimated Import Y2^	Deficit
2008	48099.84	62538.14	-14438
2009	49182.83	63942.86	-14760
2010	50265.82	65347.58	-15082
2011	51348.81	66752.3	-15403
2012	52431.79	68157.02	-15725
2013	53514.78	69561.74	-16047
2014	54597.77	70966.46	-16369
2015	55680.76	72371.18	-16690
2016	56763.75	73775.9	-17012
2017	57846.73	75180.62	-17334
2018	58929.72	76585.34	-17656
2019	60012.71	77990.06	-17977
2020	61095.7	79394.78	-18299

Composition of Indian Foreign Trade

Indian foreign trade has undergone a change in its composition over the years. In 1948-49, tea, jute manufacturers, cotton manufactures, oilseeds, hides and skins, and metals and ores constituted 71 per cent of total Indian exports. This dependence on a few commodities not only introduced an element of instability in export prospects but was bound to weaken the country's position in regard to larger questions of policy [Planning Commission, (1950)]. Import consisted of manufactured goods and food grains. On the eve of independence in 1947 ,exports consisted chiefly of raw materials and plantation crops while imports composed of light consumer goods and other manufactures [Mathur, (2006)]. The composition of trade has changed considerably. Today the manufactured goods and services dominate the export basket.

The composition of exports shows a perceptible shift in this decade i.e. 2000s from light manufactures to heavy manufactures and petroleum crude and products as shown in Table 6. The share of textiles and ready-made garments (RMG) has fallen dramatically by 11.1 percentage points in 2006-07 over 2000-01 followed by gems and jewellery, leather and leather manufactures and handicrafts. Share of engineering goods and petro products has increased by 7.6 percentage points and 10.7 percentage points, respectively. The share of primary products has declined somewhat with the decline in share of exports from agricultural and allied sector being partly offset by a rise in the share of ores and minerals by 2.8 percentage points. The share of chemicals, including petrochemicals, has increased marginally. The share of petroleum crude and products has risen further to 18 per cent in the first half of 2007-08 from 15 per cent in 2006-07. Engineering goods' share also maintained a rising trend in 2007-08. Export growth in 2006-07 was driven mainly by petroleum products with 59.3 per cent growth and engineering goods with 38.1 per cent growth. The perceptible increase in the share of petroleum products in total exports reflected not only the rise in POL prices but also India's enhanced refining capacity. The rising share of engineering goods reflected India's revival of heavy manufactures. Induced by strong international minerals, after growing at a compound annual growth rate (CAGR) of 50 per cent in the first half of this decade, moderated to 12.6 per cent in 2006-07 [Finance Ministry, (2008)].

Table 6. Commodity Composition of Exports

Commodity group	Share (per cent)					CAGR 2000-2001 to Apr-Sep	Growth rate (per cent)[a]			
				Apr-Sep					Apr-Sep	
	2000-2001	2005-2006	2006-2007	2006-2007	2007-2008	2004-2005	2005-2006	2006-2007	2006-2007	2007-2008
I. Primary products	16	15.4	15.1	13.5	13.4	16.9	18.9	19.8	18.5	16.7
Agriculture & allied	14	10.2	10.3	9.5	9.3	9	19.8	23.5	24.7	15.1

Ores & minerals	2	5.2	4.8	4	4.1	49.9	17.4	12.6	6	20.6
II. Manufactured goods	78.8	72	68.6	68.4	67.4	15.3	19.6	16.9	18.1	15.9
Textile incl. RMG	23.6	14.5	12.5	12.9	11.1	4.3	20.4	5.7	33.5	1.2
Gems & jewellery	16.6	15.1	12.6	12.7	13	16.8	12.8	2.9	-0.6	20.4
Engineering goods	15.7	20.7	23.3	22.8	23.5	25.4	23.4	38.1	48.1	21.2
Chemical & related products	10.4	11.6	11.2	11.1	10.4	21.7	17.3	19.1	28.4	10.2
Leather & leather manufactures	4.4	2.6	2.4	2.4	2.3	5.5	11.1	12.1	7.7	12.7
Handicrafts (Incl. carpet handmade)	2.8	1.2	1.1	1.1	0.8	-5.3	30.3	4.1	5.2	-14.5
III. Petroleum, crude & products (including coal)	4.3	11.5	15	16.5	17.9	38.7	66.2	59.3	106.2	27.6
Total exports	**100**	**100**	**100**	**100**	**100**	**17**	**23.4**	**22.6**	**27.3**	**17.6**

Source: Finance Ministry (2008).

The composition of imports showed much less change than that of exports as shown in Table 7. POL continues to be the single major item of import with its share stabilizing at the 30-31 per cent level. The share of capital goods imports shows the sharpest rise of about 4.9 percentage points in 2006-07 over 2000-01 due to a 3.7 percentage point rise in the share of transport equipment and 1.6 percentage point rise in the share of non-electrical machinery (excluding machine tools). It has, however, plateaued at 13 per cent in the first half of 2007-08. The greatest decline is in the import share of pearls and precious and semi-precious stones, reflecting the fall in export share of gems and jewellery. Imports of gold and silver have been at around 8 per cent though it has increased to 10 per cent in the first half of 2007-08. Share of electronic goods imports has increased to 9 per cent, while food and allied imports show a marginal fall in share due to the fall in the share of edible oils, though import of cereals has shot up in 2006-07 from a negligible level. With the rise in crude oil prices, growth in POL imports continued to be high in 2006-07 though it moderated in the first half of this fiscal [Finance Ministry, (2008)].

Table 7. Commodity composition of imports

Commodity group	**Share (per cent)**					**CAGR 2000-2001 to Apr-Sep.**	**Growth rate (per cent)[a]**			
				April-Sep.					**April-Sep.**	
	2000-2001	2005-2006	2006-2007	2006-2007	2007-2008	2004-2005	2005-2006	2006-2007	2006-2007	2007-2008
Food & allied products	3.3	2.5	2.9	2.3	2.2	24.3	-4.7	42.4	-5.8	26.6
1. Cereals	0	0	0.7	0.1	0.1	16.1	36.8	3589.6	803.8	-55.5
2. Pulses	0.2	0.4	0.5	0.3	0.5	38	41.3	53.8	9.6	92.8
3. Edible oils	2.6	1.4	1.1	1.2	1.2	17.2	-17.9	4.2	-11.8	32.9
Fuel (of which)	33.5	32.1	33.2	36.3	33.6	18.5	44.8	29	39.8	18
4. POL	31.3	29.5	30.8	33.8	31	17.5	47.3	30	41.2	16.9
Fertilizers	1.3	1.3	1.6	1.7	1.9	17.2	59.4	52.4	54.4	48.2
Capital goods (of which)	10.5	15.8	15.4	13.1	13.2	28.9	62.5	21.8	44.3	28.3
5. Machinery except electrical & machine tool	5.9	7.4	7.5	8.1	8.2	26.2	49	24.9	39.5	28.3
6. Electrical machinery	1	1	1.1	1.1	1.1	25.6	25.9	30.3	37.9	28.6
7. Transport equipment	1.4	5.9	5.1	2.1	2.5	57.7	104.2	6.8	55.7	51.2
Others (of which)	46.3	43.7	43.8	37.8	40.4	23.5	21.1	24.6	-2.8	36.4

Commodity group	Share (per cent)					CAGR 2000-2001 to Apr-Sep. 2004-2005	Growth rate (per cent)[a]			
				April-Sep.					April-Sep.	
	2000-2001	2005-2006	2006-2007	2006-2007	2007-2008		2005-2006	2006-2007	2006-2007	2007-2008
8. Chemicals	5.9	5.7	5.2	5.6	5.2	23.6	23.2	14.1	13.2	19.8
9. Pearls, precious &semi precious stones	9.6	6.1	4	4.1	4.2	18.3	-3.1	-18	-32.8	30.6
10. Gold&silver	9.3	7.6	7.9	7.7	10.3	24.5	1.5	29.4	-3.1	71
11. Electronic goods	7	8.9	8.6	9	8.9	29.9	32.5	20.6	34	26.2
Grand total	**100**	**100**	**100**	**100**	**100**	**22.2**	**33.8**	**24.5**	**23.5**	**27.7**

Source: Finance Ministry (2008).

Direction of Foreign Trade Export and Import in India

Table 8 and Table 9 show the geographic structure of Indian foreign trade. The direction of Indian foreign trade has changed in lasts few years. Over the last 60 years, India's foreign trade has undergone a complete change in terms of composition and direction [Mathur, (2006)]. Traditionally, EU and USA used to be the major trading partners of India.

Table 8. Geographic Structure of Foreign Trade of India 1995-2001

Direction of Foreign Trade Export and Import (1995-96 to 2000-01) US $ million												
Group/Country	Exports						Imports					
	95-96	96-97	97-98	98-99	99-00	00-01	95-96	96-97	97-98	98-99	99-00	00-01
I. OECD Countries	17705.1	18601.4	19484.9	19264	21106.6	23473.6	19209.2	19456.6	21335.8	21859.7	21364.3	20157.9
A. EU Of which:	8708.3	8655.3	9144.6	8946.6	9382.4	10410.8	10303.2	10624.8	10680.6	10723.8	10967.8	10510.2
1. Belgium	1120.4	1092.7	1215.5	1287.9	1367.7	1470.6	1701.9	2251.7	2668.1	2876.8	3681.3	2870
2. France	747	716.2	759.6	829.7	897.3	1020	840.7	768.1	797.7	719.6	718.2	640.8
3. Germany	1977.4	1893.1	1923.7	1851.9	1738.4	1907.6	3145.1	2831.1	2528.8	2140.7	1841.6	1759.6
4. Italy	1014	933.7	1115.2	1055	1119.8	1308.8	1064.3	987.4	921.7	1088.3	734.6	723.6
5. Netherlands	769	852.4	803.8	763.5	885.8	880.1	570.2	494	445.2	464.2	470.9	437.5
6. UK	2010.8	2046.9	2140.8	1855.4	2034.8	2298.7	1917.7	2134.7	2443.6	2621.4	2706.8	3167.9
B. North America	5825.8	6908.4	7236.1	7672.6	8973.8	9961.6	4242.6	3999.3	4137.8	4025.8	3944.2	3412.1
1. Canada	305.4	353	433.2	473	578.3	656.5	381.2	313.4	420.9	385.6	380.5	397.1
2. USA	5520.4	6555.4	6802.9	7199.6	8395.5	9305.1	3861.4	3685.9	3716.9	3640.2	3563.7	3015
C. Asia and Oceania of which:	2651.9	2456.9	2408.7	2096.2	2153	2263.6	3551.8	3584.1	3714.2	3999.1	3714.1	2984.3
1. Australia	375.7	385.4	438.3	387.4	403.3	405.9	1021.9	1317.2	1485.6	1445	1081.8	1062.8
2. Japan	2215.6	2005.9	1898.5	1652	1685.4	1794.5	2467.6	2187.4	2144.9	2465.7	2535.8	1842.2
D. Other OECD Countries Of which:	519.1	580.7	695.5	548.6	597.4	837.6	1111.6	1248.4	2803.2	3111	2738.2	3251.3
1. Switzerland	281.6	299.9	367.5	319.1	353.7	437.7	1020.5	1127.3	2640.7	2942.4	2597.7	3160.1
II. OPEC of which:	3080.8	3232.5	3535.1	3560	3902.4	4864.4	7649	10154.1	9419.4	7775.3	12864.6	2711
1. Indonesia	662.4	591.8	437.3	185.3	325.6	399.8	461.1	598.7	731.6	829.1	958.8	910.2
2. Iran	155	195	171.7	159.1	152.1	227	598.2	874.4	633	473.7	1251.3	211.2
3. Iraq	0.6	2.2	11.2	36.3	49.4	84	0	0	185.4	151.2	199.7	6.9
4. Kuwait	135.5	154.7	185.9	164.7	154.3	199.1	1970.1	2404.9	2299.5	1501.1	1912.2	112.7
5. Saudi Arabia	482.3	577.2	690.1	774.3	742.5	822.9	2024.7	2769.7	2508.3	1831.5	3016.5	621.1
6. UAE	1428.3	1476	1692.4	1867.6	2082.7	2597.5	1606.6	1736.1	1780	1721.2	2334.2	659
III. Eastern Europe of which:	1340	1098.5	1283.3	1052.9	1292.9	1317.8	1673.8	1102.7	1114.6	863.9	994.6	850.2

Direction of Foreign Trade Export and Import (1995-96 to 2000-01) US $ million)												
1. Romania	29.9	17.8	15.3	17.6	12.6	12.2	148.1	154	61.4	43.2	20.1	21.7
2. Russia	1045	811.2	953	709.4	947.9	889	856.3	628.4	678.2	545.5	623.2	517.7
IV. Developing Countries of which:	9196.6	10033	10304.4	9212	10453.4	12998.2	8140.4	8415.2	9610.8	11885.4	14510.1	11134
A. Asia	7307.8	8133.9	7972.4	6844.5	8205.5	10037.9	6426	6573.4	7258.9	8535.2	9942.2	8459.5
a) SAARC	1720.6	1701.6	1610.9	1679.2	1394.6	1928.5	256.5	241.6	234.3	465.6	397.7	465.8
1. Bangladesh	1049.1	869	786.5	995.6	636.3	935	85.9	62.2	50.8	62.4	78.2	80.4
2. Bhutan	17.2	22	13.3	9.6	7.6	1.1	34.7	33.8	13.4	6.1	18	21.1
3. Maldives	15.7	10.4	8.7	8.4	7.3	24.6	0.2	0.2	0.2	0.1	0.4	0.2
4. Nepal	160	165.7	170	122.4	151.2	140.8	49.1	64.1	95.2	144.9	188.6	255.1
5. Pakistan	76.8	157.2	143.2	106.1	92.9	186.8	45.1	36.2	44.4	214.4	68.2	64
6. Sri Lanka	401.7	477.4	489.2	437.1	499.3	640.1	41.4	45.2	30.2	37.7	44.2	45
b) Other Asian Developing Countries of which:	5587.2	6432.3	6361.5	5165.3	6810.9	8109.4	6169.4	6331.8	7024.6	8069.6	9544.5	7993.7
1. Hong Kong	1821.4	1862.6	1932	1880.6	2510.9	2640.9	388	319.1	316.3	449.3	817.9	852.1
2. South Korea	448.3	518.5	467.6	307.9	476.6	450.8	824.8	883.6	1001.8	1394.4	1273.3	893.8
3. Malaysia	393.2	531.1	489.9	321.7	447.1	608.2	902.7	1107.5	1178.9	1608.4	2024	1176.8
4. Singapore	901.6	977.5	779.7	517.5	672.7	877.1	1091.9	1063.3	1197.9	1384.2	1534.4	1463.9
5. Thailand	472.9	447.1	344	321	449.6	530.1	169.7	197.2	233.3	273.1	327.8	337.9
B. Africa of which:	1512.4	1420	1634.5	1757.2	1550.3	1951.5	1131.4	1287.7	1756.4	2626.6	3640.5	1981.6
1. Benin	11.3	16.6	20.7	27.1	28.4	45.1	15.5	9.8	13.5	12.8	42.7	52.1
2. Egypt Arab Republic	164.3	157.5	253.4	270.3	236.7	357.5	72.6	65.2	192.6	192.4	443.9	38.8
3. Kenya	245.1	168.5	123.9	145.2	116.7	140.9	15	19.6	20.9	36.2	21	19
4. Sudan	30.6	26.9	44.6	55	71.5	97.8	15.3	6.3	4.5	15.1	6.6	8
5. Tanzania	81.4	67.6	68.9	85.4	82	102	96.3	73.9	68	124.6	124.5	59.5
6. Zambia	35.2	32.1	26.6	17.3	23.3	22.5	59.9	104.1	97.2	39.1	26	11.6
C. Latin American Countries	376.4	479.1	697.5	610.2	697.6	1008.8	583.1	554.1	595.5	723.7	927.5	692.9
V. Others	18.8	22.5	45.9	40.4	37.1	76.7	3	3.8	3.9	4.4	4.4	1.6
Total Trade	**31794.9**	**33469.7**	**35006.4**	**33218.7**	**36822.4**	**44560.3**	**36675.3**	**39132.4**	**41484.5**	**42388.7**	**49670.7**	**50536.5**

Note: Exports of Petroleum Products are taken into account in total exports, but are not included in country-wise details.

Source: *Handbook of Statistics on Indian Economy*, Reserve Bank of India, 2001; IndiaStat Database (2008).

On the eve of independence in 1947, foreign trade of India was typical of a colonial and agricultural economy. Trade relations were mainly confined to Britain and other commonwealth countries [Mathur, (2006)]. However, after the Look East policy of 1990s, the share of East Asian countries in general and China in particular in recent years has increased.

Table 9. Geographic Structure of Foreign Trade of India 2001-2006

Direction of Foreign Trade Export and Import in India (2001-2002 to 2005-2006) (US $ Million)										
Group/Country	**2001-2002**		**2002-2003**		**2003-2004**		**2004-2005**		**2005-2006 (P)**	
	Exports	**Imports**	**Exports**	**Imports**	**Exports**	**Imports**	**Exports**	**Imports**	**Exports**	**Imports**
I. OECD countries	21622.1	20640.6	26382.6	23301.1	29629.2	29572.1	36494.8	39989.9	45459.7	46606.6
A. EU of which:	9845.9	10436.5	11522.5	12541.7	13890.0	14717.0	17539.6	18713.0	22222.4	22341.9
1. Belgium	1390.6	2763.0	1661.8	3711.9	1805.7	3975.9	2509.7	4588.9	2853.1	4705.4

Direction of Foreign Trade Export and Import in India (2001-2002 to 2005-2006) (US $ Million)

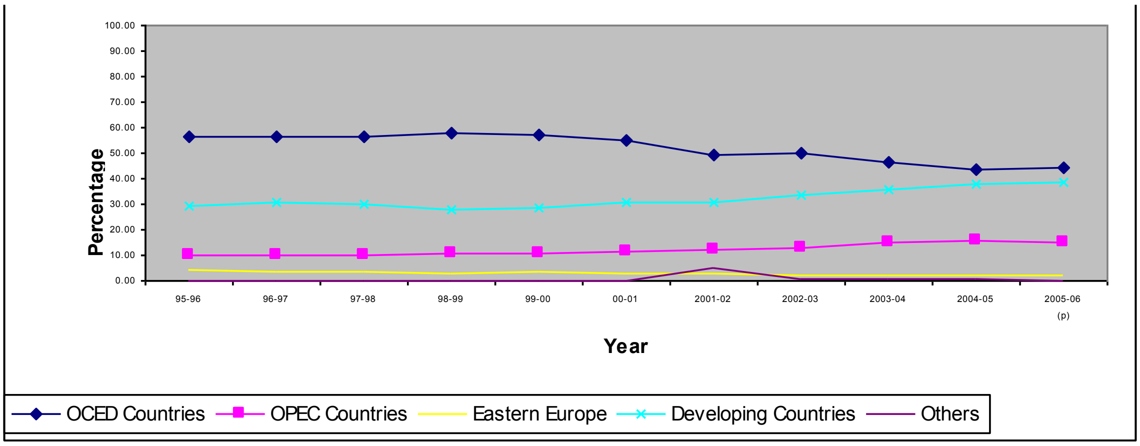
Group/Country	2001-2002		2002-2003		2003-2004		2004-2005		2005-2006 (P)	
	Exports	Imports	Exports	Imports	Exports	Imports	Exports	Imports	Exports	Imports
2. France	945.0	844.3	1074.1	1094.2	1280.9	1090.2	1680.9	1894.1	2047.6	1764.3
3. Germany	1788.4	2028.1	2106.7	2404.5	2544.6	2918.6	2826.2	4015.3	3517.0	5818.4
4. Italy	1206.5	704.8	1357.1	812.0	1729.4	1071.0	2286.0	1373.1	2490.0	1828.7
5. Netherlands	863.9	466.5	1047.9	385.7	1289.1	535.6	1604.9	791.5	2456.1	1039.1
6. UK	2160.9	2563.2	2496.4	2777.0	3023.2	3234.3	3681.1	3566.2	5146.0	3898.2
B. North America	9098.2	3679.0	11594.0	5009.9	12253.2	5760.7	14632.6	7777.1	18212.2	8673.0
1. Canada	584.8	529.4	698.3	566.3	763.2	725.9	866.8	775.7	1008.7	895.1
2. USA	8513.3	3149.6	10895.8	4443.6	11490.0	5034.8	13765.7	7001.4	17203.5	7778.0
C. Asia and Oceania of which:	1990.7	3534.7	2435.9	3249.5	2379.6	5395.8	2941.4	7187.6	3411.5	8599.5
1. Australia	418.0	1306.1	504.2	1336.8	584.3	2649.2	720.2	3824.5	811.8	4851.1
2. Japan	1510.4	2146.4	1864.0	1836.3	1709.3	2667.7	2127.9	3235.1	2458.7	3552.4
D. Other OECD countries of which:	687.3	2990.3	830.2	2500.0	1106.5	3698.6	1381.3	6312.2	1613.6	6992.2
1. Switzerland	409.1	2870.7	382.7	2329.9	449.9	3312.7	540.9	5939.9	474.1	6525.5
II. OPEC of which:	5224.5	2965.8	6884.6	3479.4	9544.4	5609.2	13207.4	10022.5	15223.1	11013.6
1. Indonesia	533.7	1036.8	826.1	1380.9	1127.2	2122.1	1332.6	2617.7	1370.5	2933.5
2. Iran	253.0	283.8	654.7	258.3	918.1	266.8	1231.4	410.2	1176.4	685.9
3. Iraq	206.8	0.0	214.9	0.0	75.2	0.1	131.2	1.1	145.4	2.1
4. Kuwait	206.2	73.7	250.6	179.5	319.1	142.5	421.4	305.9	507.7	460.5
5. Saudi Arabia	826.4	464.0	940.7	504.7	1123.3	737.8	1412.1	1301.2	1806.9	1617.4
6. UAE	2491.8	915.1	3327.5	957.0	5125.6	2059.8	7347.9	4641.1	8592.8	4312.1
III. Eastern Europe of which:	1254.8	946.8	1248.1	1139.9	1555.4	1628.9	1780.2	2514.2	1960.3	3689.5
1. Romania	11.4	48.4	27.4	45.6	47.8	71.9	106.0	168.4	83.4	251.2
2. Russia	798.2	535.5	704.0	592.6	713.8	959.6	631.3	1322.7	729.9	1992.0
IV. Developing countries of which:	13535.5	12776.4	17862.3	15688.2	22784.3	20567.2	31597.1	28604.2	39785.1	36808.8
A. Asia	10332.7	9264.7	13981.0	11303.8	18426.7	16269.8	24968.4	22581.3	30961.1	29849.1
(a) SAARC	2026.0	571.5	2724.1	512.0	4148.1	668.8	4440.7	950.2	5358.5	1338.7
1. Bangladesh	1002.2	59.1	1176.0	62.1	1740.7	77.6	1631.1	59.4	1632.4	118.8
2. Bhutan	7.6	23.9	39.0	32.2	89.5	52.4	84.6	71.0	99.1	88.9
3. Maldives	26.9	0.4	31.6	0.3	42.3	0.4	47.6	0.6	67.2	2.0
4. Nepal	214.5	355.9	350.4	281.8	669.4	286.0	743.1	345.8	859.4	380.0
5. Pakistan	144.0	64.8	206.2	44.8	286.9	57.6	521.1	95.0	681.9	177.5
6. Sri Lanka	630.9	67.4	921.0	90.8	1319.2	194.7	1413.2	378.4	2018.5	571.7
(b) Other Asian developing countries of which:	8306.6	8693.2	11256.9	10791.8	14278.6	15601.0	20527.7	21631.1	25602.6	28510.4
1. China, People's Republic of	952.0	2036.4	1975.5	2792.0	2955.1	4053.2	5615.9	7098.0	6721.2	10739.5
2. Hong Kong	2366.4	728.9	2613.3	972.6	3261.8	1492.7	3691.8	1730.1	4457.2	2167.6
3. South Korea	471.4	1141.4	644.9	1522.0	764.9	2829.2	1041.7	3508.8	1819.0	4343.4
4. Malaysia	773.7	1133.5	749.4	1465.4	892.8	2046.6	1084.1	2299.0	1151.6	2388.6
5. Singapore	972.3	1304.1	1421.6	1434.8	2124.8	2085.4	4000.6	2651.4	5569.8	3230.3
6. Thailand	633.1	423.1	711.2	379.0	831.7	609.1	901.4	865.9	1062.3	1201.7
B. Africa of which:	2260.9	2502.4	2575.7	3348.2	3094.4	3103.9	4478.6	3930.4	5809.2	4548.7
1. Benin	55.4	43.1	64.4	38.2	52.7	53.2	47.1	79.8	114.5	77.1
2. Egypt, Arab Republic of	462.7	99.9	298.2	226.6	367.5	98.2	444.7	152.6	664.3	218.0
3. Kenya	156.0	31.9	203.6	33.5	229.5	41.9	426.6	46.7	569.1	48.2
4. Sudan	122.2	13.4	105.2	24.5	107.4	31.2	317.4	22.9	313.5	31.7
5. Tanzania	90.8	76.1	115.5	91.2	175.8	109.3	173.9	131.7	239.9	119.4
6. Zambia	25.7	13.5	31.0	14.4	39.9	18.6	50.4	23.0	64.7	40.3

Direction of Foreign Trade Export and Import in India (2001-2002 to 2005-2006) (US $ Million)										
Group/Country	2001-2002 Exports	2001-2002 Imports	2002-2003 Exports	2002-2003 Imports	2003-2004 Exports	2003-2004 Imports	2004-2005 Exports	2004-2005 Imports	2005-2006 (P) Exports	2005-2006 (P) Imports
C. Latin American countries	941.9	1009.3	1305.6	1036.2	1263.2	1193.6	2150.1	2092.5	3014.8	2411.0
V. Others/Unspecified	2189.8	14083.7	341.8	17803.5	329.3	20771.7	456.5	30386.7	296.8	44297.9
Total trade	43826.7	51413.3	52719.4	61412.1	63842.6	78149.1	83535.9	111517.4	102725.1	142416.3

Abbr.: P – Provisional.
Source: Reserve Bank of India; IndiaStat Database(2008).

Further, the Indian foreign trade geographic structure in exports and imports in terms of percentage shows the increasing share of developing countries and decreasing share of OECD group. The geographic structure of Indian exports in percentage terms (excluding petroleum exports) is shown in Figure1.

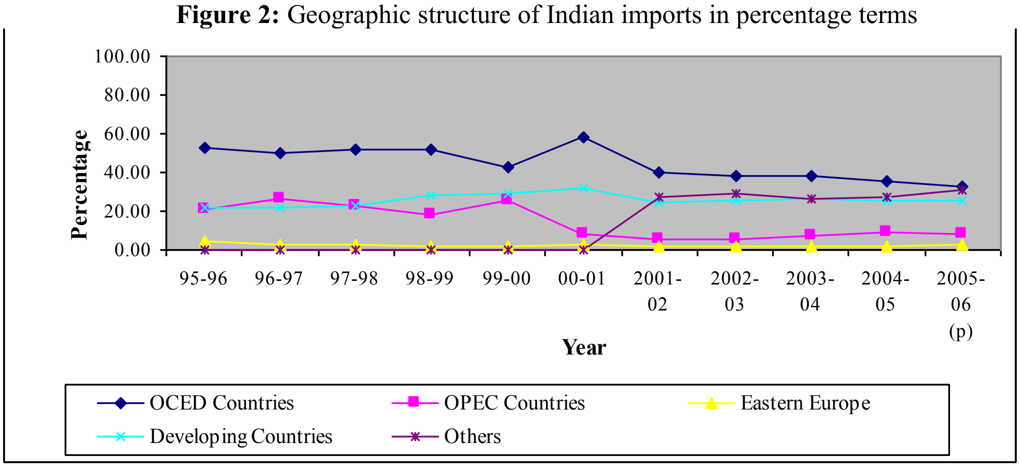
Figure1: Geographic structure of Indian exports in percentage terms



* excluding petroleum exports.

Graph 2 displays the geographic structure of Indian imports in percentage terms (excluding petroleum products), see the Figure 2.

Figure 2: Geographic structure of Indian imports in percentage terms



3. Conclusion

Indian foreign trade has progressed a lot over the last sixty years since Independence. The period can be divided into three sub-periods of 1950-1970, 1971-1991 and post-1991.The trade has stagnated and India lost its market share to other countries in 1950s and 1960s.The government policies and dominant views of import substitution and export pessimism has a negative impact. The situation improved in 1970s and exports has finally picked up in post liberalization era in general and after 2002 in particular. In terms of composition, now it is dominated by manufactured goods and services. Services exports contribution has grown rapidly in recent past. India services exports share in global exports is more than double of that of Indian manufacturing exports In terms of direction, it is now more distributed around the world and the share of East Asian countries his on rise in overall trade. Looking at the large size of the economy, the high growth rates and small share in world trade; with the help of economic theories, we can safely conclude that there is huge untapped potential for Indian foreign trade in years to come.

4. References

[1] Bhagwati, J.; Srinivasan, T.N., (1975), *Foreign Trade Regimes and Economic Development: India*, New York: Columbia University Press.

[2] Commerce Ministry, (2008), India Trade Data Bank, Ministry website. http://commerce.nic.in/eidb/default.asp (Accessed on 18th March, 2008)

[3] IBEF, (2005), Why service exports could overtake merchandise trade by 2007,IBEF website. http://ibef.org/artdisplay.aspx?cat_id=391&art_id=7340 (Accessed on Dec 18, 2007)

[4] Indiastat Database (2008), Foreign Trade, Indiastat Database website.

http://www.indiastat.com/india/ShowData.asp?secid=331671&ptid=107&level=3 (Accessed on 18th March, 2008)

[5] Joshi, V.; Little, I.M.D., (1994), India*: Macroeconomics and Political Economy*, *1964-1991*, Washington DC: World Bank and Oxford University Press.

[6] Kathuria, S., (1996), Export Incentives: The Impact of Recent Policy Changes in India, *Indian Economic Review*, Vol 31, No.1, pp. 109 – 26.

[7] Malik, J.K., (2005), India's Exports- Policy defeating Exchange Rate Arithmetic, *Economic and Political Weekly*, Dec 24-31.pp 5486 – 5496.

[8] Mathur, V., (2006), *Foreign Trade of India 1947-2007 : Trends, Policies and Prospects*, New Delhi: New Century Pub.

[9] Ministry of Finance, (2008), Economic Survey 2007-08, Ministry website. http://indiabudget.nic.in/es2007-08/chapt2008/chap62.pdf (Accessed on 23rd March 2008).

[10] Nayyar, D., (1976), *India's Exports and Export Policies in the Sixties*, Cambridge: Cambridge Universtiy Press.

[11] Planning Commission, (1950) 1st Five Year Plan, Planning Commission website. http://planningcommission.nic.in/plans/planrel/fiveyr/1st/1planch30.html Accessed on 23rd March 2008).

[12] Singh, M., (1964), *India's Export Trends and the Prospects for Selfsustained Growth*, Oxford: Clarendon Press.

[13] Singh, S., (2007), Services exports soar to put India at No.10, Livemint, Nov 27. http://www.livemint.com/2007/11/22003411/Services-exports-soar-to-put-I.html

[14] Veeramani, C., (2007), Sources of India's Exports Growth in Pre- and Post- Reform Periods, *Economic and Political Weekly*, July 23-30.pp. 2419 – 2427.

[15] Virmani, A., (2003), *India's External Reforms- Modest Globalisation, Significant Gains*, Economic and Political Weekly, Vol.38, No.32, pp.3373 – 3390

FINANCIAL AND LEGAL CONSTRAINTS TO FIRM GROWTH: THE CASE OF ITALY

Domenico **SARNO**
Department of Law and Economics
Seconda Università di Napoli, **ITALY**
domenico.sarno@unina2.it

Abstract:

The aim of this study is to confirm empirically the implications of the theory about the law-finance-growth nexus. In order to verify the predictions of the theory, a panel data including three different types of data is used. All the data are referred to Italian provinces. The empirical analysis shows that between firms' growth and financial development there is a first-order relationship, while between firms' growth and legal enforcement as measured by the efficiency of the judicial system there is a second-order relationship.

Keywords: enforcement, judicial efficiency, financial development, firm's growth

JEL classification: G2, K4

1. Introduction

Differences between long-run growth rates are theoretically explained with the degree of the development of financial markets and institutions. Although there is not agreement about the causal direction exiting between the financial development and the economic growth, there is a massive body of theoretical and empirical studies showing that the growth is strongly connected with the financial development through a first order relationship. Particularly, both the ideas that financial development accelerates economic growth and that the degree of maturity of financial markets represents a good predictor of the potential of economic growth are largely accepted. This is because the financial institutions gathering the information mitigate the problems that arise from the information asymmetries and reduce the transaction costs, mobilize the private savings and improve the allocation efficiency [King and Levine, (1993); Levine, (1997)].

An related field of analysis has been growing recently; it aims to identify the causes of the differences between national financial structures. Within this approach differences between national legal systems are explained with reference of the major legal traditions: the *British Common Law* and the *French Civil Law* [La Porta *et al*., (1997), (1998); Demirgüç-Kunt and Maksimovic, (1998), (1999)]. According to this theory, the countries have developed over the years legal systems insuring the protection of the private rights in different degrees. This protection is higher in the Common Law tradition compared to the Civil Law tradition. "*The British Common law evolved to protect private property owners against the crown. This facilitated the ability of private property owners to transact confidently, with positive repercussions on financial development. In contrast, the French Civil law was constructed to eliminate the role of a corrupt judiciary, solidify state power, and restrain the courts from interfering with state policy. Over time, state dominance produced a legal tradition that focuses more on the rights of the state and less on the rights of individual investors than the British Common law.*" [Beck *et al*., (2003), pp.138 – 139).

However, beyond differences in origin, the protection of the investor rights also depends on law enforcement. The level of the enforcement can vary within the same legal systems. It is determined by the efficiency of the judicial system and its performance can significantly differ within the same national legal system because of the misallocation of resources or local imbalances between the judicial demand and supply.

The rules protecting creditor interests and contractual enforcement affect more directly the efficiency of the financial markets and thus have repercussion on the growth. The finance theory helps us to explain the microeconomics of this process more precisely. The default probability is reflected by agency costs; the larger the degree of opacity in the relationship between the firm and the financial markets, the higher the agency costs. Those costs rise more if there is a lack of enforcement; as a consequence the access to external financing sources becomes more difficult and credit rationing and request for collaterals are increasing. Therefore, the firm will be able to take market opportunities to grow only if it can produce internal resource flows devoted to finance it. That makes financial constraints more likely; economic growth, therefore, becomes more volatile due to this higher dependence of firm growth on internal finance.

In this work I propose a preliminary analysis of this theoretical field through a intra-country analysis utilising a microeconomic data set. More precisely, the paper aims to verify empirically the existence of a first-order relationship between firm growth and financial development and of a second order relationship between firm growth and judicial efficiency. The data set consists of a closed sample of Italian SMEs and a number of variables measuring financial development and judicial efficiency in Italian provinces.

The paper is organised as follows. In the next section I discuss the theoretical foundations of the law-finance-growth relationship. Section 3 presents the characteristics of the data set. In section 4 an preliminary, graphical analysis is proposed, while in section 5 I illustrate the empirical model and the estimates obtained. Some concluding remarks are reported in the last section.

2. Financial development, legal system and firm growth: a brief review

The finance theory allows us to understand how market imperfections restrict the possibility for the firm to use the external resources in order to finance its investment projects. According to this theory, market imperfections feed the conflict between the firm's insiders and the outsider investors; this conflict originates from, on the one hand, information asymmetries and, on the other hand, the lack of the contractual enforcement.

Information asymmetries make it impossible or too costly to sign complete contracts between lenders and borrowers; inefficient enforcement makes default profitable. The size of this negative effects is determined by the degree of the development of the financial and legal systems: they are the less extensive the more developed are financial and legal institutions [Beck *et al*., (2005); Demirgüç-Kunt and Maksimovic, (1998), (1999); La Porta *et al*., (1997), (1998), (2000); Levine, (1998), (1999)].

Growth theory has definitively rejected the idea that the financial systems adjust to real growth passively; instead, the financial systems are seen as determinant to the process of economic growth. The main function of financial institutions and markets is to mitigate the problems arising from information imperfections. Particularly, the financial institutions are dedicated to gathering information; this makes the *ex-ante* acquisition of information cheaper and makes the *ex post* monitoring of the firm's behaviour more efficient. The delegated monitoring function minimizes transaction costs and improves resource allocation; moreover, it affects capital accumulation, technological innovation and long-run growth dynamic [Diamond, (1984); Bernanke and Gertler, (1990); Levine (1997)][18].

Financial institutions play an important role in reducing the opacity in the relationship between firms and financial markets and in facilitating firms' access to external finance. Of course, it is not easy to empirically verify how much well behaved financial institutions resolve the problems deriving from the information imperfections; nevertheless, the analysis using microeconomic data set help us to evaluate their importance. On this regard the theory suggests that, if the costs related to acquisition of the information by the outsider investors increase, the difficulties for the firm to get access of external resources increase as well. Empirical works show that the firms receive more external funding where financial markets are more developed [for example, Beck *et al*., (2005)].

Information imperfections increase the wedge between the external and the internal finance; this makes internal resources cheaper as primary financial source of funding. As a consequence, the

[18] The large literature on the finance-growth nexus dates back to the paper by Gurley-Show, Goldsmith and McKinnon. The recent research is reviewed by Levine (1997).

growth of the firm is more strictly dependent on the internal resources that the firm can produce and can devote to finance itself[19,20].

The literature on the financial constraints to the growth of the firm draws on these theoretical discussions. There is now a wide body of research in this field, producing many important but controversial results. This literature is heavily focused on the problem of the relationship between capital structure and the growth of the firm. The theoretical basis are the Modigliani-Miller propositions that establish the independence of the growth of the firm from capital structure and, on the opposite side, the pecking order hypothesis affirming that external finance is not a perfect substitute for internal funding and, consequently, the growth is not independent from the internal finance[21].

Empirical studies have tried to identify a positive relationship between firm's growth and internal finance. In particular, there have been questions as to whether the sensitivity of investments to cash flow is a good indicator of the presence of financial constraints to growth [Fazzari *et al*., (1988); Bond, Meghir, (1994); Hubbard, (1998)][22,23].

The sources of agency conflicts lie in the capital structure of the firm. The property right theory helps us to identify these sources. According to this theory, different funding sources are associated with different kinds of contracts defining the allocation of property rights. Those contracts minimize the transaction costs and simultaneously act in protection of investor's interests. For example, an equity issuance causes an *ex ante* reallocation of shares of the property rights within the firm; here investor's protection lies in the owners' behaviour. By contrast, in the case of debt issuance this protection is ensured through an *ex post* transfer of property rights to the firm's assets; the creditor can recover totally or partially his credit if the firm defaults [Williamson, (1985), (1988); Aghion and Bolton, (1992); Hart, (1995)].

Within this framework the problems arise when the mechanism of the transfer of the property rights on firm's assets cannot adequately protect creditor's interests when the firm acts to reduce the value of the transferred assets[24].

Therefore, the function of the judicial system in the lender/borrower relationship is to force the repayment of debt upon the borrower who would not do it spontaneously. Indeed, the borrower could find non-repayment profitable when the benefits of the default exceed the perceived costs of the penalty. Since the weakness of the enforcement reduces the credibility of the threat and the costs related to it, the inefficiency of the judicial system tends to increase the opportunistic behaviour of the borrowers. As a consequence, lenders discount the probability of such event by reducing credit supply and making firm's access to the external finance more difficult.

This argument is proposed by Jappelli *et al*. (2005). In this work trough a theoretical model the authors show that judicial inefficiency increases incentives for borrowers to behave in an opportunistic way with no relation whatsoever to the probability of a successful outcome for investment projects. This happens because the borrower can always deny the repayment of its debt to the lender leading it to a trial. The advantages of default for the borrower increase if: **a.** the borrower can appropriate of a

---

[19] Information imperfections cause the credit rationing. Collaterals exist to reduce the effects of the credit rationing. On the one hand, this permits to supply credit to small opaque firms; on the other hand, it is a device that allows controlling managers' behaviour.

[20] It is believed that the relationship banking can reduce the negative impact of information imperfections. Close relationships between firms and banks facilitate the information transfer and improve the relationship of the firms with the financial markets. Nevertheless, the empirical results are contrasted. On the one hand, relationship banking gives benefits with regard to the decreasing interest rate and increasing credit availability. On the other hand, it may carry high private and social costs arise as banks exploit market power deriving from its information advantage.

[21] The two propositions of MM are contained in the 1958' and 1961' articles [Modigliani and Miller, (1958), (1961)] and have been subsequently extended by Miller [Miller, (1977)]; the pecking order hypothesis dates back to the works of Myers and Majluf [Myers and Majluf, (1984); Myers, (1984)].

[22] The references cited in the text are representative of a massive body of theoretical and empirical studies analysing the problems related to the relationship between the capital structure and the growth of the firm and about to investment financing; the cited work of Hubbard (1997) contains a broad review of this literature.

[23] The thesis that the sensitivity of investment on cash flow can be interpreted as financial constraints has been famously put into question by Kaplan and Zingales [Kaplan and Zingales, (1997), (2000)]. They believe that the sensitivity of investment to cash flow indicates profitable opportunities by the firm.

[24] In the finance theory those strategies are seen as costly because they decrease the firm value.

part of the income arising from the investment or can consume part of the collateral; **b.** the cost of the trial is sustained entirely by the lender. Within the model the hypothesis about agents' behaviour is developed assuming that the credit markets are competitive or, alternatively, imperfect; the conclusions of this research suggest that the improvement of the enforcement reducing the opportunistic behaviour leads to an increase of the credit supply (or a decrease of credit rationing) and a decrease of the collaterals[25].

3. Characteristics of the data set

The analysis covers the 103 Italian provinces and the years from 1995 to 2003 years. The data set collects three different types of data: microeconomic data on firms, data related both to financial development and to the efficiency of the judicial system.

The microeconomic data on firms comes from a closed sample of Italian manufacturing SMEs. This sample has been extracted from the last three surveys on the Italian Manufacturing Sector performed by Capitalia, and previously by Mediocredito Centrale, for the three three-years 1995-1997, 1998-2000 and 2001-2003[26]. It is composed of 533 small and medium sized firms (<250 employees); the small firms (<50 employees) represent 89,1% of the sample (475 firms). A size variable is determined on sample's data: **1.** total assets (ASSET) of the firms.

The data on the financial system come from Bank of Italy statistics. They consist of: **a.** the number of the bank branches by province; **b.** the distribution of the loans granted to residents by province. In order to define the variables used for the estimate data from ISTAT source on population and GDP by province are in addition utilised. The model then uses two proxies for financial development: **2.** the number of bank branches for 100,000 inhabitants (BRANCHS_POP) and **3.** the loans/GDP ratio (LOANS_GDP) by province.

The data related to judicial system are obtained from the ISTAT statistics of the civil justice. Those statistics include the annual flow of trials and the stock of pending trials; moreover the statistics contain several measures of efficiency of the legal system. Since 2000 this data is broken down at the court district level, corresponding to provinces or below; for previous years the data are available at the level of appeal's court districts, which encompass one or more provinces[27]. The data by province are obtained by matching of the court data and appeal's court data and provinces. Two proxies of the efficiency of civil judicial system are defined: **4.** the ratio of closed trials on total trials (CLOSEDTRIALS) and **5.** the mean length of trials (LENGHTRIALS); the first one is an efficiency index, the second one an inefficiency measure.

Finally, the data set used for the estimate has been obtained by matching of the firm-level data and the values of financial and judicial variables of the province where the firm has its headquarters [28].

4. The efficiency of the italian judicial system and the relationship with the financial system

With regard to the empirical analysis it is decisive to consider the efficiency of the judicial system. There is much empirical work devoted to these issues [Marchesi, (1998), (2001), (2003); ISTAT-MIPA, (2006)]. The national judicial statistics contains standard efficiency indicators related to the trials at various judiciary levels. However the subject is very complex.

Generally civil contentious can be classified into four classes: property and obligations, succession, work and family. The procedures are divided into three classes of trials: cognitive, precautionary and executive. The cognitive trial is finalised to the assessment of the judicial

[25] The effects on the interest rates are ambiguous because they are affected by the market structure. In a monopolistic market the lender can appropriate of the borrower surplus entirely; therefore, an improvement in the judicial efficiency causes an increase in interest rates. In competitive markets, the rate rises if there is inside collateral, while if there is outside collateral the effect on the rates is ambiguous.

[26] The *Survey of Italian manufacturing firms* draws on a representative sample of the population of manufacturing firms; it is based on questionnaires filled by the sampled firms and of the data sheet for the most of them.

[27] There are 164 court districts and 26 appeals court districts.

[28] The sample contains firms belonging to 92 of the 103 Italian provinces.

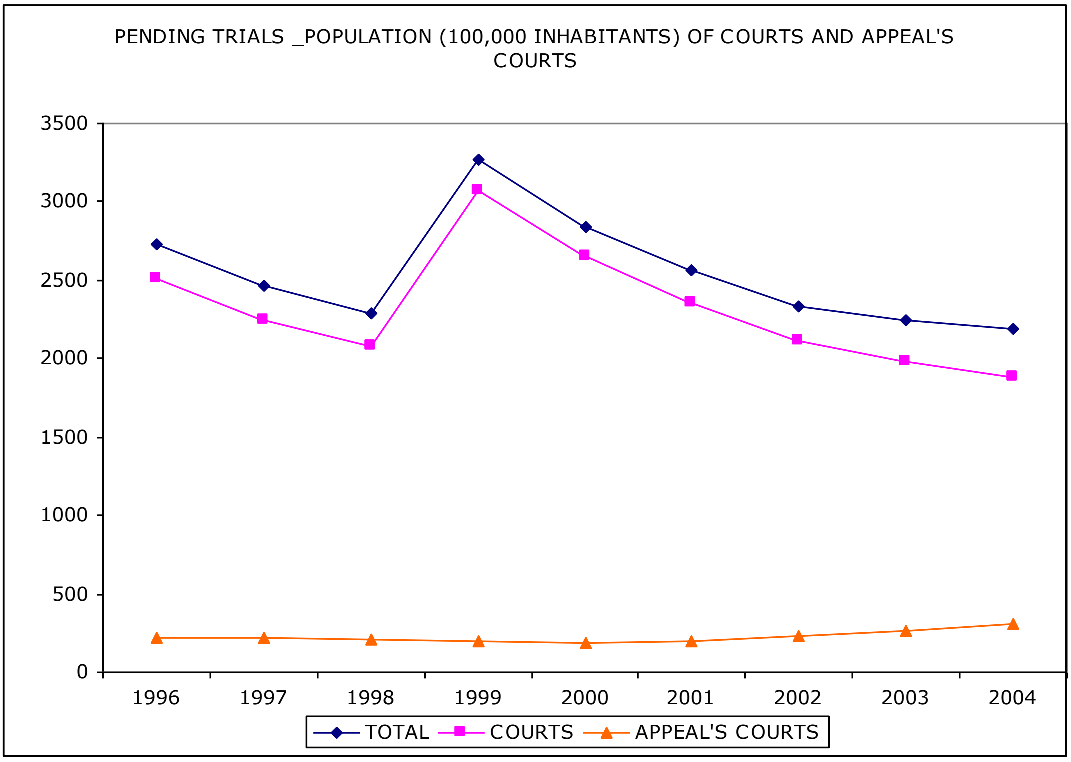
controversy while the other two classes of trials precede or following to it[29]. The efficiency analysis can be restricted to the cognitive trials. They concern all the different degrees of justice and they are concentrated in the courts[30]. Statistics are available for all the various degrees; they relate to the ingoing and outgoing trials and the stock of pending trials; moreover there are several indicators scaled to population and efficiency indexes[31,32].

The trend of two efficiency indexes is taken here as an example. The graphs in the Figure 1 depicts the trend in the years 1996-2006 of 1a) the number of the I° and II° degree trials pending both at first-degree courts and at the courts of the appeals scaled for the population (100,000 inhabitants) and 1b)

Figure 1: Stock of pending trials and mean lenght of trials – Italy

The graphs depict **a.** PENDING TRIALS scaled of the population (100,000 inhabitants) opened in the courts and the appeal's courts and **b.** LENGTH OF TRIALS (days) in the 1996-2004 years. Data are absolute value.

Source: Processing on National Institute of Statistics (ISTAT)



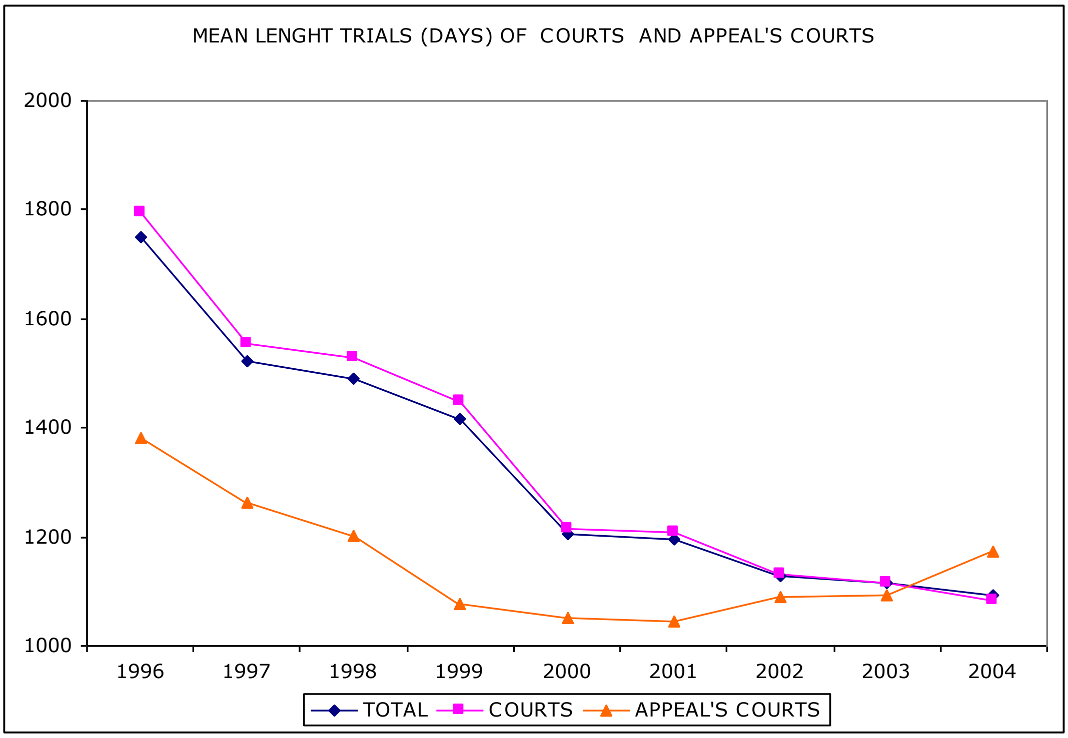
a.

[29] More precisely, the precautionary trial precedes and complements the cognitive trial because it should guarantee the successful of the cognitive and executive trials. By contrast, the executive trial makes the execution of the sentence and therefore follows the cognitive trial.

[30] Since 1998 a significant reorganization of the Italian judicial system has been implemented. The court has become the only one judicial office for the I° degree trials. The less important trials in civil law are delegated to "peace officer". Therefore the three different degree of the civil justice are: the peace officer or the court, the court of appeals and, finally, the Cassation court.

[31] The efficiency indexes available from national statistics are: the turnover ratio, the closed trials ratio, the index of controversial settlement and the mean length of the trials. A more exhaustive definition of the indexes used in this work is provided in Appendix 1.

[32] Cognitive trial concern the general contentious of the civil justice. Marchesi (2003) finds that in years 1975-1998 about the 60% of annual cognitive trials involved economic controversies.

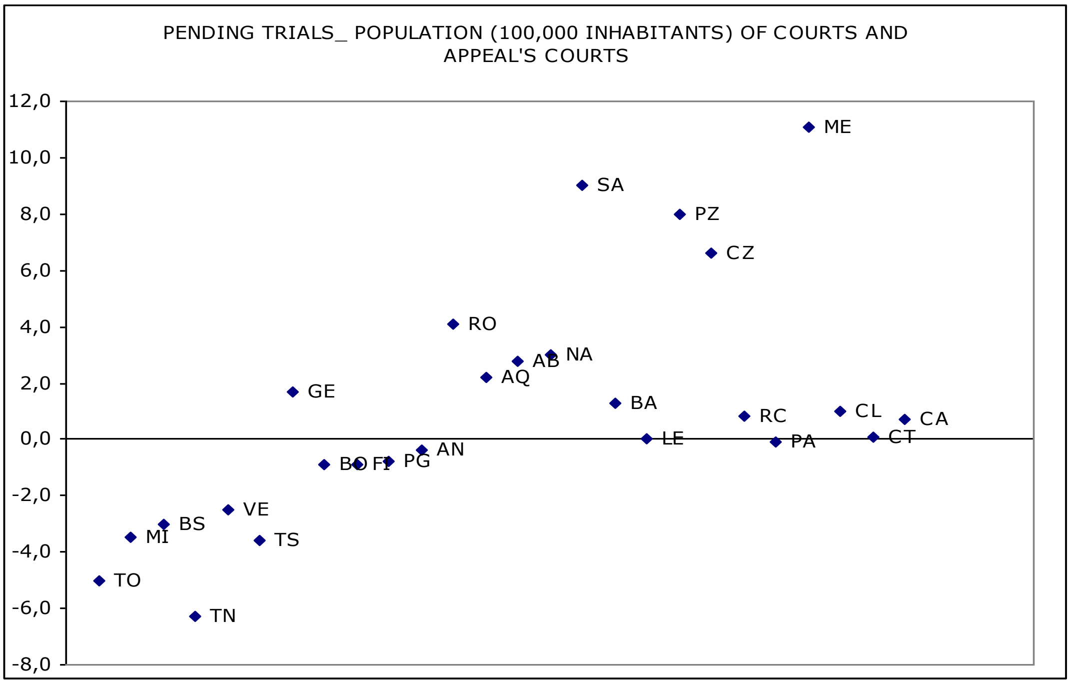


b.

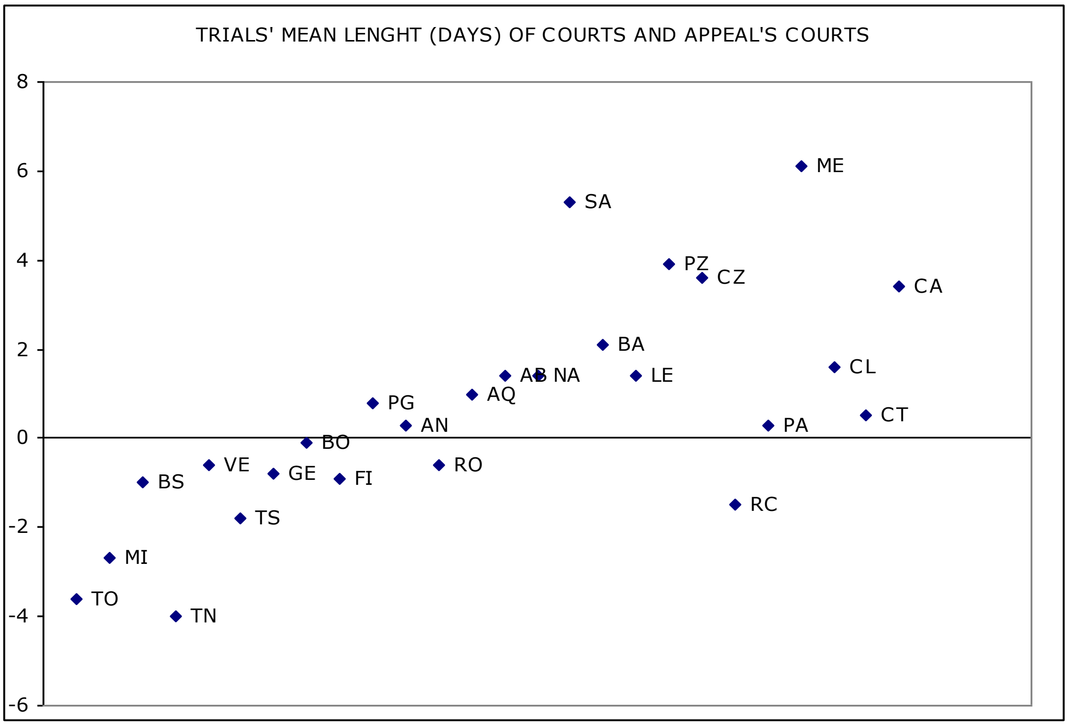
Figure 2. Stock of pending trials and mean lenght of trials by judicial district

The graphs depict **a.** PENDING TRIALS scaled of the population (100,000 inhabitants) opened in the courts and the appeal's courts and **b.** LENGTH OF TRIALS (days) in the 1996-2004 years by judicial districts. Data are expressed as residuals on the mean value. The legenda of the judicial districts is in Appendix 1.

Source: Processing on National Institute of Statistics (ISTAT)



a.



b.

The trend of the pending trials shows a weak improvement of the efficiency, in particular since 1999. The mean length of trials is significantly decreasing. In both of the cases, the improvement have almost exclusively concerned I° degree trials.

Figure 2 reports the territorial differences in the two previous efficiency indexes of the judicial system. The graphs show the distribution of the residuals with regard to the national mean of 2a) the pending trials and 2b) the length of the trials of the judicial district and of appeal's courts[33]. Since the graphs reflects the regional distribution of the judicial districts from the Northern to Southern Italian regions moving towards right, it appears clearly that the judicial activities in the northern judicial districts are more efficient than in Southern judicial districts.

Before presenting the estimates, it can be useful at this stage of the analysis to verify through an descriptive, graphical analysis the relationship between the development of the financial system and the efficiency of the judicial system.

In Figure 3 the two pairs of graph represent the relationships between the two proxies of the financial development by province and the two proxies of the judicial efficiency by province. The data express the mean of the period and then are transformed into index numbers around a national mean value (=100). The graphs include trend lines as well.

The pair of graphs 3a) and 3b) shows an increasing relationship between the two financial indicators and the rate of closed trials; the other pair, 3c) and 3d), shows an decreasing relationship between the two financial variables and the mean length of trials.

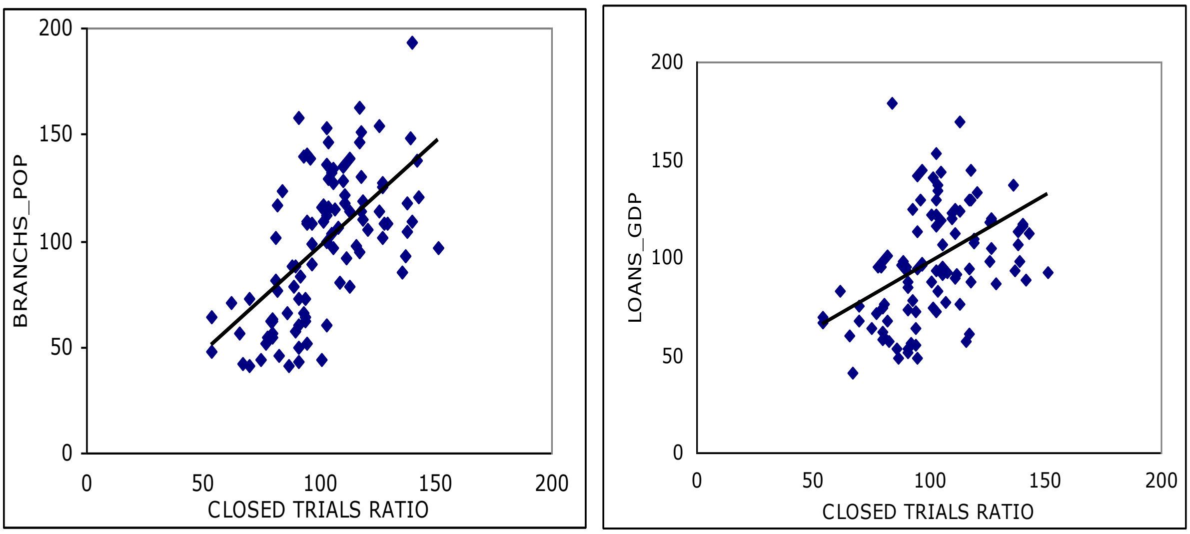
In all cases, it appears clearly that there is a relevant positive relationship between enforcement as measured by the efficiency of the judicial system and the development of the financial markets.

5. Financial and legal constraints to firm growth: the econometric analysis

The previous model can be represented by two equations: in the first, the growth of the firm is dependent by the financial development; in the next, the financial development is dependent by the judicial efficiency.

[33] As we have mentioned above, there are 26 courts of appeal's districts; moreover, there are 3 judicial sections that are not included in the graphs. The districts are in accordance with the judicial demand: they are fewer in the Northern and Central regions of Italy, while they are more numerous in the South and in the Islands.

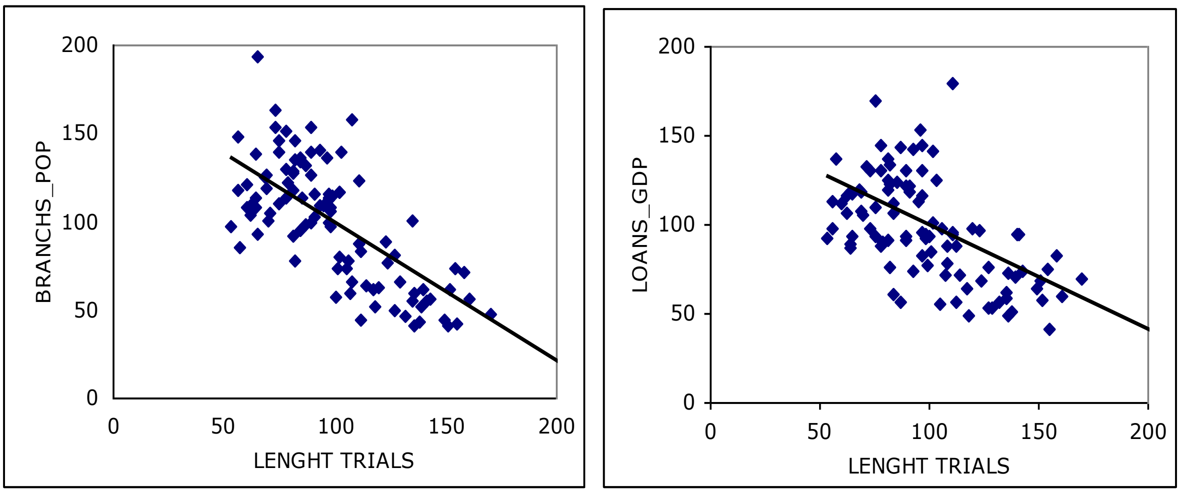
Figure 3. Financial development and judicial efficiency relationship



a. **b.**

The graphs contain the provincial number index (Italy=100). On the y-axis there are the provincial number index of the financial development variables (BRANCH_POP (100,000 inhabitants) and the LOANS / VALUE ADDED RATIO), on the x-axis there are the provincial number index of the judicial efficiency variables (CLOSED TRIALS RATIO and LENGTH OF TRIALS).

Source: processing on Banca d'Italia and ISTAT data.



c. **d.**

Therefore, the panel representing the reduced form of the model can be represented as

$$log\ Y_{i,t} = \beta_1\ log\ Y_{i,t-1} + \beta_2 F_{it} + \beta_3\ G_{it} + \alpha_{i,t} + \eta_i + \varepsilon_{it}$$

where firm growth is $log\ Y_t$-$log\ Y_{t-1}$ and Y the size variable of the firms [assets (ASSET)], F represents the two variables measuring financial development [branch banks/population ratio (BRANCHS_POP) and loans/GDP ratio (LOANS_GDP)] and, finally, with G as the two proxies of the judicial efficiency [closed trials/total trials ratio (CLOSEDTRIALS) and mean length of the trials (LENGHTRIALS)]; α and η are the specific and temporal effects, ε the error term.

All the variables are expressed as natural logarithm; the statistics and the correlation matrix of those variables are presented in Table 1.

The estimates are obtained through the DPD methodology proposed by Arellano-Bond (1994). It is known that in the Arellano-Bond estimate the first differences of the variables are used to eliminate the fixed effects; subsequent GMM estimates are performed using the independent lagged variables as instruments. The choice of the instruments is a crucial issue; it can be seen that, due to the correlation between the lagged dependent variable and the error term, the more suitable instruments are the two lag independent variables[34].

The estimates are in Table 2[35]. The Sargan tests reject the null hypothesis about the restrictions of over-identification, while the AR tests confirm the presence of the first-order correlation and reject the presence of second-order autocorrelation between the residuals; the test F is always significant.

The estimates are coherent with the theory. The coefficient related to the lagged independent variable are about 0.50/0.60 values. The positive relationship between firm growth and the proxies of the financial development are confirmed: the BRANCHS_POP coefficients are higher (0.17/0.18) compared to the LOANS_GDP coefficients (0.05/0.07). The sign of the coefficients related to the judicial system are positive for the efficiency index (CLOSEDTRIALS) and negative for the inefficiency index (LENGHTRIALS); all the coefficient values are in the range 0.20/0.24.

Table 1. Statistics and correlation matrix

Absolute values				
Variable	Mean	Std.Dev.	Minimum	Maximum
ASSETS	33535	138924	366	2778020
$ASSETS_{-1}$	31756	132775	328	2676513
BRANCH_POP	56,3	12,8	16	103
LOANS_GDP	76,4	30,5	10	169
CLOSEDTRIALS	31,2	6,6	13	58
LENGHTRIALS	991,9	301,2	348	2492
Logarithmic values				
Variable	Mean	Std.Dev.	Minimum	Maximum
ASSETS	9,165	1,291	5,904	14,837
$ASSETS_{-1}$	9,114	1,284	5,793	14,801
BRANCH_POP	3,998	0,265	2,773	4,634
LOANS_GDP	4,246	0,457	2,302	5,129
CLOSEDTRIALS	3,416	0,217	2,565	4,061
LENGHTRIALS	6,855	0,299	5,852	7,822

Correlation matrix						
	ASSETS	$ASSETS_{-1}$	BRANCH_POP	LOANS_GDP	CLOSEDTRIALS	LENGHTRIALS
ASSETS	1,000	0,992	0,052	0,104	0,049	-0,072
$ASSETS_{-1}$	0,992	1,000	0,048	0,113	0,058	-0,082
BRANCH_POP	0,052	0,048	1,000	0,234	0,282	-0,368
LOANS_GDP	0,104	0,113	0,234	1,000	0,237	-0,410
CLOSEDTRIALS	0,049	0,058	0,282	0,237	1,000	-0,892
LENGHTRIALS	-0,072	-0,082	-0,368	-0,410	-0,892	1,000

In the table are riported the statistics and the correlation matrix of the variables used for the estimates. ASSETS = Totale assets at year t; $ASSETS_{-1}$= Total asset at year t-1; EMPLOYEES = Employees al year t; $EMPLOYEES_{-1}$ = Employees al year t-1; BRANCH_POP = Branch bank / population ratio by provinces (100,000 inhabitants); LOANS_GDP = Loans / GDP ratio by provinces; CLOSEDTRIALS = Closed trials / Total Trials ratio; LENGHTRIALS = Length of trials (days).

---

[34] The instruments are obtained using the orthogonal condition between the lagged dependent variable and the errors. In order to illustrate this approach a model AR(1) can be considered $y_{it}= \rho\, y_{it-1} + \nu_{it}$ where $\nu_{it}=\mu_i + \varepsilon_{it}$

where both μ_i and ε_{it} are normally distributed with zero mean and constant variance. In order to obtain an consistent estimator by ρ for fixed T (small) and $N\to\infty$ the first order differences are calculated and the fixed effects μ_i are eliminated. The result is $\Delta y_i = \rho\Delta y_{it-1} + \Delta\varepsilon_{it}$

where y_{it-1} is correlated with ε_{t-1}. This means that if the observations begin in year t = 1, the first year where the previous relationship can be observed is t=3.

[35] We have produced estimates using the employees as dependent variable. This estimate confirms the casual relationship, but they show serial correlation between instrument with two, three and four lags and error terms. This violates the conditions imposed in the PDD approach.

Table 2. Law-finance-growth estimates

Dependent varaible: assets				
$ASSETS_{-1}$	.581	.536	.580	.514
	[.031]***	[.033]***	[.031]***	[.036]***
BRANCHS_POP	.187	.174		
	[.072]***	[.067]***		
LOAN_GDP			.074	.053
			[.011]***	[.010]***
CLOSEDTRIALS	.202		.245	
	[.040]***		[.033]***	
LENGHTRIALS		-.209		-.219
		[.031]***		[.027]***
F TEST	952.4***	1074.5***	1001.4***	1095.2***
SARGAN	300.6***	300.4***	324.9***	350.4***
AR(1)	-15.4***	-12.5***	-14.6***	-12.1***
AR(2)	-0.4	-0.3	-0.4	-0.3
N° INSTRUMENTS	84	84	84	84
N° OBSERVATIONS	3720	3720	3730	3730

This table contains the estimates of the relationship between the firm growth, the financial development and the judicial efficiency. The variables are expressed in natural logarithm. The dependent variables are: ASSETS (Total asset); the dependent variables are $ASSETS_{-1}$ and BRANCHS_POP (branch / population ratio), LOANS_GDP (Loans on GDP ratio), CLOSEDTRIALS (Closed trials / Total trials ratio) and LENGHTRIALS (length of trials in days).

Estimates: Arellano-Bond estimates (GMM estimates on the first order differences and instruments y_{t-2} e y_{t-3}); years 1995-2003. The standard deviation are in quadratic brackets; *, **, *** significance levels at 1, 5 and 10%.

SARGAN = test of sovra-identification (critical value at 1% = 25.2); AR(1) and AR(2) – Arellano-Bond test (H0 = absence of autocorrelations of first order residuals and of second order residuals, respectively).

6. Conclusions

The results of the previous analysis confirm the implications of the theory about law-finance-growth nexus. This study differs with existing works on the same subject in two ways. First, the analysis presented here is not cross-country but is, instead, performed within the same national context and focuses on cross-regional (provincial) differences in the contractual enforcement as measured by the efficiency variables of the civil justice. Secondly, microeconomic data, not macroeconomic data, is used for the empirical test.

The econometric test has confirmed that there is a first order relationship between the degree of financial development and economic growth; at the same time, the test has shown that there is a positive relationship between the growth of the firm and the efficiency of the civil judicial system.

These conclusions are in line with the basic tenets of the theory exposed above. The potential growth of firms is affected by financial development in that more developed financial markets improve the firm's access to external funding sources. This enables firms to finance their investment projects using external finance, thus mitigating the financial constraints arising when economic growth is excessively sensitive to internal finance. Financial development is enhanced if the rule system devoted to protect the investor well-behaves. The lender/borrower contract is typically incomplete because it occurs in a environment where there are information imperfections; which frequently lead to opportunistic behaviours and defaults. The improvement of the enforcement can reduce the opportunistic behaviours if the legal system efficiently enforces the rules protecting the creditors' interests and makes penalty credible in case of default.

References

[1] Aghion P., Bolton P., (1992), *An Incomplete Contracts Approach to Financial Contracting*, Review of Economic Studies, vol. 59, pp. 473 – 494.

[2] Beck T., Demirgüç-Kunt A., Levine R., (2003), *Law, Endowments and Finance*, Journal of Financial Economics, vol. 70, pp. 137 – 181.

[3] Beck T., Demirgüç-Kunt A., Levine R., Maksimovic V., (2000), *Financial Structure and Economic Development: Firm, Industry and Country Evidence*, World Bank, Policy Research Working Paper, n. 2423.

[4] Beck T., Demirgüç-Kunt A., Levine R., Maksimovic V., (2005), *Financial and Legal Constraints to Growth: Does Firm Size Matter?*, in: *Journal of Finance*, vol. 60, pp. 137 – 177.

[5] Bernanke B, Gertler M., (1990), *Financial Fragility and Economic Performance*, in: *Quarterly Journal of Economics*, vol. 103, pp. 87 – 114.

[6] Besanko D., Thakor A.V. (1987) – *Collateral and Rationing: Sorting Equilibria in Monopolistic and Competitive Markets*, in: *International Economic Review*, vol. 28, pp. 671 – 689.

[7] Bond S., Meghir C., (1994), *Dynamic Investment Models and the Firm's Financial Policy*, in: *Review of Economic Studies*, vol. 61, pp. 197 – 222.

[8] Demirgüç-Kunt A., Maksimovic V., (1998), *Law, Finance and Economic Growth*, in: *Journal of Finance*, vol. 53, pp. 2107 – 2137.

[9] Demirgüç-Kunt A., Maksimovic V., (1999), *Institutions, Financial Markets and Firm Debt Maturity*, in: *Journal of Financial Economics*, vol. 54, pp. 295 – 336.

[10]Diamond D., (984), *Financial Intermediation and Delegated Monitoring*, Review of Economic Studies, vol. 51, pp. 393 – 414.

[11]Fazzari M.S., Hubbard G.R., Petersen B.C., (1988), *Financing Constraints and Capital Investment*, in: *Brooking Papers of Economic Activity*, vol. 1, pp. 141 – 195.

[12]Generali A., Gobbi G., (1996), *Il recupero dei crediti: costi, tempi e comportamenti delle banche*, Bank of Italy: Temi di discussione n. 265.

[13]Hart O., (1995), *Firms, Contracts and Financial Structure*, Oxford University Press, Oxford.

[14]Hubbard G.R., (1998), *Capital-Market Imperfections and Investment*, in: *Journal of Economic Literature*, vol. 36, pp.193 – 225.

[15]Istat-Consorzio Mepa, (2006), *Ritardi della giustizia e ricadute sul sistema economico*, Roma.

[16]Jappelli T., Pagano M., Bianco M., (2005), *Courts and Banks: Effects of Judicial Enforcement on Credit Markets*, in: *Journal of Money, Credit and Banking*, vol. 37, pp. 223 – 244.

[17]Kaplan S.N., Zingales L., (1997), *Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints*, in: *Quarterly Journal of Economics*, vol. 112, pp. 169 – 215.

[18]Kaplan S.N., Zingales L., (2000), *Investment-Cash Flow Sensitivities are not Measures of Financial Constraints*, in: *Quarterly Journal of Economics*, vol. 115, pp. 707 – 712.

[19]King R.G., Levine R., (1993), *Finance and Growth. Schumpeter Might Be Right*, in: *Quarterly Journal of Economics*, vol. 108, pp. 713 – 737.

[20]La Porta R., Lopez-de-Silanes F., Shleifer A., Vishny R.W., (1997), *Legal Determinants of External Finance*, in: *Journal of Finance*, vol. 52, pp. 1131 – 1150.

[21]La Porta R., Lopez-de-Silanes F., Shleifer A., Vishny R.W., (1998), *Law and Finance*, in: *Journal of Political Economy*, vol. 106, pp. 1113 – 1155.

[22]La Porta R., Lopez-de-Silanes F., Shleifer A., Vishny R.W., (2000), *Investor Protection and Corporate Governance*, in: *Journal of Financial Economics*, vol. 58, pp. 3 – 27.

[23]Levine R., (1997), *Financial Development ad Economic Growth: Views and Agenda*, in: *Journal of Economic Literature*, vol. 35, pp. 688 – 726.

[24]Levine R., (1998), *The Legal Environment, Banks and The Long-run Economic Growth*, in: *Journal of Money, Credit and Banking*, vol. 30, pp. 596 – 620.

[25]Levine R., (1999), *Law, Finance and Economic Growth*, in: *Journal of Financial Intermediation*, vol. 8, pp. 8 – 35.

[26]Marchesi D., (1998), *L'inefficienza della giustizia civile: conseguenze sull'economia*, in: Cassese S., Galli G. (a cura di), *L'Italia da semplificare. Le istituzioni*, Bologna, Il Mulino.

[27]Marchesi D., (2001), *Giustizia: tempi e interazioni con il sistema economico*, in: ISAE, Rapporto Trimestrale, n.5, pp.1 – 36.

[28]Marchesi D., (2003), *Litiganti, avvocati e magistrati. Diritto ed economia del processo civile*, Bologna, Il Mulino.

[29]Miller M., (1977), *Debt and Taxes*, in: *Journal of Finance*, vol. 32, pp.261 – 275.

[30]Modigliani F., Miller M., (1958), *The Cost of Capital, Corporation Finance and the Theory of Investment*, in: *American Economic Review*, vol. 48, pp. 261 – 297.

[31]Modigliani F., Miller M., (1961), *Dividend Policy, Growth and the Valuation of Shares*, in: *Journal of Business*, vol. 34, pp. 411 – 433.

[32]Myers S.C., (1984), *The Capital Structure Puzzle*, in: *Journal of Finance*, vol. 39, pp. 575 – 592.

[33]Myers S.C., Najluf N.S., (1984), *Corporate Financing and Investment Decisions when Firms have Informations that Investors do not have*, in: *Journal of Financial Economics*, vol. 13, pp. 187 – 221.

[34]Petersen M., Rajan R.G., (1994), *The Benefits of Lending Relationships: Evidence from Small Business Data*, in: *Journal of Finance*, vol. 49, pp. 3 – 37.

[35]Sharpe S.A., (1990), *Asymmetric Information, Bank Lending and Implicit Contracts: A Stilized Model of Customer Relationship*, in: *Journal of Finance*, vol. 45, pp. 1069 – 1087.

[36]Stiglitz J., Weiss A., (1981), *Credit Rationing in Markets with Imperfect Information*, in: *American Economic Review*, vol. 71, pp. 67 – 127.

[37]Williamson O., (1985), *The Economic Institutions of Capitalism*, New York, The Free Press.

[38]Williamson O., (1988), *Corporate Finance and Corporate Governance*, in: *Journal of Finance*, vol. 43, pp. 567 – 591.

Appendix 1. Data by province

The data are referred to the 103 Italian provinces for the period 1995-2003. They are classified for three types: microeconomic data of the firms, data related to the financial system and data related to the judicial system (civil justice).

- **Firm's microeconomic data** – The data are available from a closed sample formed by 533 manufacturing firms. The sample is extracted from the last three surveys of the Italian manufacturing firms on three three-years periods: 1995-1997, 1998-2000, 2001-2003.

Source: Capitalia, Medio Credito Centrale

- **Financial system data** – The data cover the distribution of the bank branches by province and the distribution of the loans to the residents by province.

Source: Bank of Italy

- **Judicial system data** – The data refer to the trials requiring adjudication of substantive rights (cognitive trials). They concern a broad aggregate of all cases on civil matters, among which credit and commercial matters, which represents about 60% of the total trials. This data are available for court district since year 2000 and for appeal's court district in previous years.

Source: Italian National Institute of Statistics (ISTAT).

In order to determine the set of variables utilized for the estimates, the data of the GDP and of the Populations by province for the period 1995-2003 are considered.

Source: Italian National Institute of Statistics (ISTAT).

The variables are:

1. ASSET – Total assets extracted from the data sheet of the firms of the closed sample,
2. BRANCHS_POP – Branch banks / Population ratio (100,000 inhabitants) by provinces
3. LOANS_GDP – Loans / GDP ratio by provinces
4. CLOSEDTRIALS – Closed trials / Total Trials ratio.

The index is calculated as: Closed Trials / (Initial Pending Trials + Occurring trials)

5. LENGHTRIALS – Mean length of the trial (days).

The index is calculated as:

(Initial Pending Trials + Final Pending Trials) / (Occurring Trial + Closed Trials) x 365

Legenda of judicial districts

TO	Torino	NA	Napoli
MI	Milano	SA	Salerno
BS	Brescia	BA	Bari
TN	Trento	LE	Lecce
VE	Venezia	PZ	Potenza
TS	Trieste	CZ	Catanzaro
GE	Genova	RC	Reggio di Calabria
BO	Bologna	PA	Palermo
FI	Firenze	ME	Messina
PG	Perugia	CL	Caltanissetta
AN	Ancona	CT	Catania
RO	Roma	CA	Cagliari
AQ	L'Aquila		
AB	Campobasso		

ANALYSIS OF HIGH FREQUENCY DATA ON THE WARSAW STOCK EXCHANGE IN THE CONTEXT OF EFFICIENT MARKET HYPOTHESIS

Paweł **STRAWIŃSKI**, Robert **ŚLEPACZUK**
University of Warsaw, **POLAND**
pstrawinski@wne.uw.edu.pl rslepaczuk@wne.uw.edu.pl

Abstract:

This paper focuses on one of the heavily tested issue in the contemporary finance, i.e. efficient market hypothesis (EMH). However, we try to find the answers to some fundamental questions basing on the analysis of high frequency (HF) data from the Warsaw Stock Exchange (WSE). We estimate model on daily and 5-minute data for WIG20 index futures trying to verify daily and hourly effects. After implementing the base methodology for such testing, additionally we take into account the results of regression with weights, i.e. robust regression is used that assigns the higher weight the better behaved observations. Our results indicate that we observe the day of the week effect and hour of the day effect in polish data. What is more important is the existence of strong open jump effect for all days except Wednesday and positive day effect for Monday. Considering the hour of the day effect we observe positive, persistent and significant open jump effect and the end of session effect. Aforementioned results confirm our initial hypothesis that Polish stock market is not efficient in the information sense.

Keywords: high-frequency financial data, robust analysis, pre-weighting, efficient market hypothesis, calendar effects, intra-day effects, the open jump effect, the end of session effect, emerging markets.

JEL Classification: G14, G15, C61, C22

1. Introduction

The article investigates market inefficiencies on the Warsaw Stock Exchange, hereafter (WSE). We provide one of the first high frequency data analyses for European emerging market describing widely known phenomena of day of the week effect and hour of the day effect. The analysis is conducted using both high frequency data (5-minute returns for the period: 2003-2008) and daily data (for 10 years time span: 1998-2008) for WIG20 index futures[36]. We estimated separate models for 5-minute and daily log returns. To remove sudden price jumps that normally dominate the model, data filtering methods are used and provide result robustness. Financial return data are characterized by phenomena such as asymmetry, leptokurtosis and autocorrelation. This departure from normality has a significant influence on modelling strategy. We include regression with and without weights, additionally showing the results for each year separately (for analysis on HF data). To overcome existing problems we use pre-weighting and rely rather on the medians than the means to make our findings more robust in statistical sense.

The direct reason for our research was the need to answer some basic questions concerning EMH basing the research process on HF data. There exist extensive studies concerning EMH, however the analysis are performed on daily data mainly. Our aim is to bridge this gap and shed some light on what happen during trading day. After analysis of financial literature, not only focusing on the issue of EMH, we hypothesized that Polish stock market should still reveal some patterns of inefficiency especially connected with daily volatility patterns [Slepaczuk and Zakrzewski, (2008)], which clearly disclose strong differences between different parts of the stock market session. Basing on the presumptions built on these patterns we formulated the following hypothesis concerning Polish stock market:

- There exists intra-day effects, especially revealed in the open and final jump;
- The existence of the day effect should die out in the consecutive years of the sample being tested, as a market gradually changes from emerging to developed;

[36] We use index futures data in order to stress increasing liquidity of this type of instrument and almost 10 times lower costs of transaction. It is of special importance when we consider practical implementation of EMH research's results.

▪ The robust methodology will enable us to reveal some patterns of distribution, which were not visible in the process of standard regression. Intra-day data are prone to fractional integration and quasi-outliers problems. By quasi-outliers we treat natural observations, that are a result of sudden change of market environment that would be treated as outlier in standard methodology which were further tested in this research. Additionally, taking into account that Polish futures market (standardized products listed on the stock exchange) is the biggest within ten new member states, we assume that interpretation of our results could be extended to other emerging European markets.

The structure of our paper is as follows. After introduction in the first section, next one contains short description of the contemporary inefficiencies revealed in the process of research being conducted on the emerging market data. We focused only on the papers which introduced the concept of EMH in the early 60 and 70. [Fama, (1965), (1970); Samuelson, (1965)], defined well recognized inefficiencies [Rozeff and Kinney, (1976), Lakonishok and Maberly, (1990) etc.] and finally which recently tested the hypothesis on the emerging market data [Nath and Dalvi, (2004), Das and Arora, (2005)]. Next section contains the detailed description of the methodology and data with special attention to the empirical strategy used in the process of estimation. Our results, divided into daily and HF, and with and without weights, are presented in the fourth section. The summary of our research with short references to the future research is presented in the last section.

2. Market inefficiencies: Day and Intra-day effects

The Efficient market hypothesis is closely connected with Eugene Fama and his seminal paper in 1965, where he identified three forms of efficiency in the information sense (week, semi-strong and strong form). The EMH simply states that, basing on which form of efficiency do we choose, specified set of information is included in the prices of listed assets and then we can not create any strategy beating the market. Historically, the first research which introduced the concept of efficient market hypothesis was dissertation of Luis Bachelier (1900) largely ignored until Samuelson (1965) who popularised Bachelier's work among economists.

The EMH started to gain acceptance between economists researching this subject from 60. and even until now it is one of the most heavily researched issue in contemporary finance. Through all this years researchers found many significant anomalies of capital market, no matter if we mean weak, semi-strong or strong form, which contradict the validation of initially defined hypothesis.

Focusing on the EMH in the weak form stating that asset prices should incorporate the information contained in the historical time series we could define following market anomalies:

▪ The calendar effects: the month-of-the-year (the January effect – Rozeff and Kinney (1976), Haugen and Lakonishok (1988), the week-of-the-month [Ariel, (1987), Lakonishok and Smidt, (1988)], the day-of-the-week [the Weekend effect – Cross, (1973), Lakonishok and Maberly, (1990), the Holiday effect], and the hour-of-the-day [Smirlock and Starks, (1986), Harris, (1986), Wood, McInish and Ord (1985)]- the open and final jump effect).

▪ Correlation of rates of return in the short term – significant and positive relation [Campbell, Lo and MacKinlay, (1997)];

▪ Correlation of rates of return in the long term – significant and negative coefficients [Poterba and Summers, (1988), Fama and French, (1988)];

▪ Contrarian Strategies – long and short term [Lehman, (1990)];

▪ Momentum strategies – mid term approach [Jegadesh and Titman, (1993)].

Evidently these are only the best documented anomalies but they gave the arguments for opponents of the EMH and enabled its reconsideration on the ground of new results. Above mentioned anomalies were especially documented while testing emerging market data what contributed to the formulation of hypothesis that they are characteristics of the market which are not fully developed and should disappear in the later phase of development. However, searching the results of the emerging market researches we can not fully agree with this notion.

Latest researches on emerging market data as well as developed ones confirm the existence of day-of-the-week effect: Nath and Dalvi (2004), Lyroudi *et al.* (2002), Das and Arora (2005), Bhattacharya *et al.* (2003), which most often result in significant Monday and Friday. When we consider hour-of-the-day effect we observe the open jump effect [Harvey and Huang, (1991)] and final jump effect [Guin, (2005)] which are closely connected with the daily patterns of volatility but there is much less evidence concerning this issue.

Taking into account researches on Polish data we have to admit that although we have four extensive works focusing on the EMH in the information sense in all forms [Czekaj *et al*. (2001), Szyszka (2003), Jajuga (2000), Buczek, (2005)] the authors did not pay special attention to calendar effects. Only in Szyszka (2003) week-of-the-month effect and day-of-the-week are verified and some patterns of inefficiency are found: significant positive Monday and significant positive first week of the month. The papers mentioned above tested only the data on the daily level what naturally made impossible to reveal hour-of-the-day effects. That was one of the reasons why we conducted this research on Polish HF data trying to reveal the calendar effects widely identified in the literature, and compare them with overall notion that emerging markets most often reveal some patterns of inefficiency.

3. Research methodology and data description

Our empirical analysis is based on high-frequency financial data for WIG20 index futures. We based our study on the continuous time series for futures, where expiring futures contract was replaced by the next series, where the number of open positions achieves the higher value. It is one of the most common ways of creating continuous time series for futures. We do not have enough data for the longer period of time because of the short time to expiration of individual future contract, therefore we had to create continuous futures index.

WIG20 consists of 20 largest companies quoted on WSE and is computed as a weighted measure of the prices of its components. The daily data span is from 2nd February 1998[37] to 31st of March 2008. Unfortunately, the 5-minute data, which were supplied by Information Products Section from WSE, are available from 2nd June 2003 to 31st of March 2008. The number of 5-minute returns for a trading day depends on the trading hours for futures contracts but this have been changed once during our research period. The trading was from 9:00 to 16:00 for the time period from 2nd June 2003 until 30th September 2005 and from 9:00 until 16:30[38] for the next two years from 3rd October 2005 until 31st of March 31 2008. Thus, we had 84 or 90 five-minute returns for a day in the research period. All returns were computed as the first difference in the regularly time-spaced log prices of WIG20 index futures, with the overnight return included in the first intraday return. After correction for outliers (three on the basis of five-minute intervals) we get a total of 2547 trading days and a total of 103122 five-minute intervals.

Table 1: The descriptive statistics for log returns of analysed index futures returns for the period from 2nd June 2003 to 31st March 2008.

Statistics	Daily data	5-minute data
Mean	.0003047	.0008837
Standard Deviation	.0182882	.1601220
Variance	.0003345	.0256390
Skewness	.0922582	-.4723653
Kurthosis	5.283044	38.98765
N	2547	103122

Source: own computation based on WSE data

Table 1 summarizes the data descriptive statistics[39]. In the left column are daily data characteristics and in the right we have five-minute data. Five minutes average return are slightly higher than daily returns, however both numbers are not significantly different from zero. Analysing both returns series, we can evidently see higher than normal kurtosis and small skewness. The daily returns are slightly skewed. On the other hand, five-minute interval returns are left skewed. The

[37] WIG20 was quoted from 20th January 1998, however there was a very few transactions. Continuous data are avaliable since February 1998.

[38] In practice, the continuous trading finished at 16:10, then the close price was settled between 16:10 and 16:20, and next investors could trade only on the basis of close price until 16:30. Therefore, we could even say that we have 86 instead of 90 intervals in the second period.

[39] Visual presentation of daily data in Appendix.

distribution of the returns is leptokurtic, i.e. is almost symmetric and has fat tails and a substantial peak at mean value. Testing for normality, we get the same results for both data sets, i.e. the statistics reveal non-normality of the data sets tested[40]. Analyzing mean, standard deviation, kurtosis and skewness we observe some patterns of distribution (Table 2 and Figure 1 and 2).

Table 2. The descriptive statistics calculated separately for each half an hour interval during the day.[a]

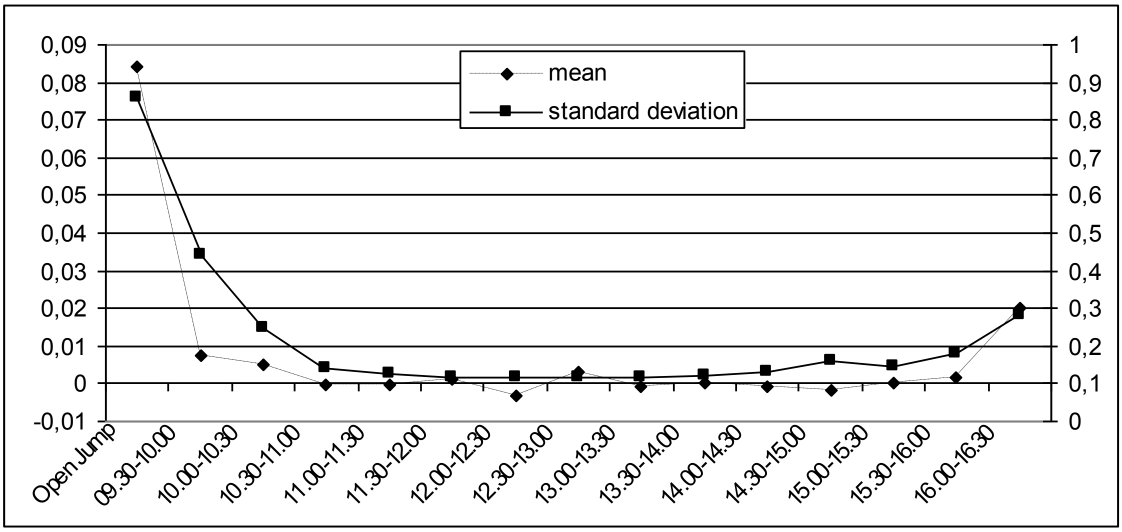
Period	Return	Std dev	Skewness	Kurtosis
Open Jump	0.0845649	0.7296813	-0.8220053	6.574713
09.00-09.30	0.0147039	0.3208124	-1.0734150	28.388360
09.30-10.00	-0.0031634	0.1418127	0.5770158	17.932900
10.00-10.30	-0.0023202	0.1642937	-0.1403808	7.934801
10.30-11.00	-0.0001690	0.1384769	0.1073290	7.559757
11.00-11.30	-0.0002199	0.1248630	-0.1761427	8.627885
11.30-12.00	0.0014125	0.1172726	0.1155043	8.827273
12.00-12.30	-0.0030406	0.1143765	0.2130371	11.481020
12.30-13.00	0.0029511	0.1149602	0.0132158	10.408220
13.00-13.30	-0.0007801	0.1148532	-0.1983475	8.339746
13.30-14.00	0.0004350	0.1208250	-0.0525039	12.341730
14.00-14.30	-0.0009016	0.1306344	0.6896316	31.747580
14.30-15.00	-0.0014733	0.1608028	-0.2442220	10.006480
15.00-15.30	0.0001727	0.1464368	-0.1424887	8.962907
15.30-16.00	0.0014616	0.1804251	0.2183429	9.870681
16.00-16.30	0.0201590	0.2812534	0.2936624	6.279024

[a] The descriptive statistics calculated for Wig20 index futures on the basis of 5-minute data in the analyzed period.

Source: own computation based on WSE data

J-shape of the fluctuations of series in Figure 1, presenting standard deviation and mean during the day and significant differences in kurtosis and skewness suggest us that some trading periods during the day could be more important than the others. We decided to check our presumption in the process of formal analysis described in the next sections.

Figure 1: The fluctuations of mean and standard deviation during the stock market session. [a]

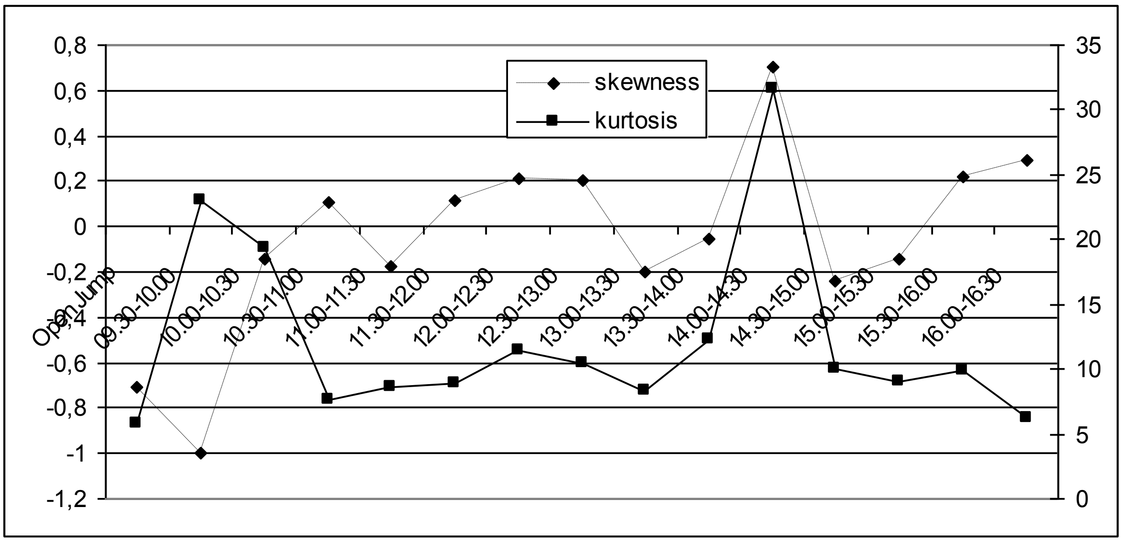


[a] The descriptive statistics calculated for WIG20 index futures on the basis of 5-minute data in the analysed period

Source: own computation based on WSE data

[40] Normality tests are not presented to conserve space. They are available upon request.

Figure 2: The fluctuations of kurtosis and skewness during the stock market session. [a]



[a] The descriptive statistics calculated for WIG20 index futures on the basis of 5-minute data in the analysed period.

Source: own computation based on WSE data.

Actual market data frequently contain a small fraction of unusual data points which are not consistent with Gaussian assumption. This problem is especially severe in a case of financial data, where, for instance, returns are well-known from being leptokurtic and fat-tailed. When normality of the error term cannot be assumed, OLS estimate will allow for unbiased estimation only for linear function of dependent variable. Moreover, for distribution with outliers, or heavy-tailed distribution statistical properties of estimators are problematic [Koenker, Portnoy (1987)]. In the literature some *ad-hoc* remedies are proposed such as removing or down-weighting questionable data. Chang and Lakonishok (1992) showed that robust estimation methods for returns data provide better estimates.

Robust statistical methods provide an attractive alternative and have recently attracted growing attention. Such estimators give less weight to "outlier" observations that poorly fits to the data. A M-estimation is a generalization of classical inference in econometrics. The M estimator is obtained by minimizing

$$\psi(t)F(t) \tag{1}$$

where $\psi(t)$ is a weighting function and F(t) is a distribution function.

Two popular weighting functions are used in applied researches: Huber psi, and Tukey bisquare. The former assigns median value for outlying observations. The latter yields more efficient estimator for heavy tailed distribution [Kleiner et al (1979)].

$$\psi(t) = \begin{cases} t(1-t^2)^2 & |t| < c \\ 0 & |t| > c \end{cases} \tag{2}$$

Tukey bisquare function for *c=3* is similar to rule of a thumb that advocates to drop observations which are more than 3 standard deviations away from the centre of data. The actual value of *c* should be determined by size of empirical sample and distance from normality.

From the available information about value of the future contract, intervals on which market did not operate, were excluded (i.e. weekends and national holidays). As a result we have 2547 working days in the sample. In the further analysis we use log-returns instead of returns to remove the market expansion effect. Returns are calculated as a difference between successive log values of index:

$$r_t = \ln(P_t) - \ln(P_{t-1}) \tag{3}$$

for daily return, P_t is a FW20 close value at day *t*, and P_{t-1} is a FW20 close value at day *t-1*, respectively. For 5-minute return P_t is FW20 close value at the end of time interval *t*, and P_{t-1} is the same value one time period before. For the cases in which there was no data in previous time interval or information was not available or not reliable due to know market perturbations we used values form last available data period to calculate the adequate returns. Further in the analysis we have included OPEN dummy that captures the effect of market opening. Usually the close value of the index differs from opening value on the following day.

$$r_t = \sum_{1}^{q} \alpha r_{t-q} + \beta * D_j + \gamma OPEN_t + \varepsilon \qquad (4)$$

t denotes current value of the variable and *t-q* value lagged for q periods. For daily data, lagged value is the value from previous trading day, for 5-minute data is the value for previous trading period. The lagged dependent variables are used for several reasons. Firstly, they capture persistency of return patterns on the market. Secondly, they are used to remove potential autocorrelation. It is widely known in the literature that high-frequency data are characterized by long correlation patterns, especially when an underlying variable is fractionally integrated. Proposed model has no constant term. Instead we use set of dummy variables. D_j are dummies for day or half-hours, and OPEN is a dummy that captures the difference between close value of the index on previous day and the opening value. Its role is to capture changes that happen during closure of the market.

To estimate model (4) we apply a robust estimator from a class of M-estimators and rely completely on robust methodology. For selection of threshold values of *c*, that detects deviant observations we use robust analogy of 99,5% confidence interval[41]. We decided to use quite wide interval due to having lot of observations. However, we are aware that it is hard to distinguish outliers from the values that occur with low probability in large samples. As a result, we excluded from our sample only extremely deviant observations by setting their weights to zero. For the remaining observations, we used scaled residuals, i.e. residuals divided by mean absolute deviation, as input for weighting function and then calculated weights. Application of ψ-function is analogous to weighted estimation. The primary role is to eliminate the influence of outliers on the estimate. This procedure assigns higher weights to those observations with relatively small residuals, and lower weights for those with large ones. The next section contains the results of our estimations separately for the model on daily and 5-minute data.

4. Results for daily and HF data

The analysis on daily data is not an aim of the research. The investigation is performed for benchmark purpose only but the main attention will be paid to the results on HF data presented further in this section.

Model on daily data

The dependent variable in the analysis is a daily log-return from future contract. As explanatory variables are used lagged by 1 and 2 periods log-returns of future contract to control for market fluctuations and set of day-dummies to capture daily effects.

Table 3. Models on daily data.

Variable	Standard		Robust	
Lagged 1 period return	-.0050546		-.0217709	
Lagged 2 period return	.0486470	**		
Monday	-.0001580		-.0003544	
Tuesday	-.0005757		-.0009403	*
Wednesday	-.0006993		-.0007059	
Thursday	.0011882		.0012453	**
Friday	.0015180	*	.0011537	**

[41] The robust counterpart for 99.5% confidence interval is $MED(y) \pm 5.2(1+\frac{4}{n})MAD(y)$. Where MED(y) is a median of variable, MAD(y) stands for mean absolute deviation form the median, and *n* is a sample size.

*** 1%; ** 5%; * 10%. Significance level
Source: own computation based on WSE data

Inclusion of the second lag of dependent variable is enough to remove correlation of the error term. We have also tried to include additional lags to check the robustness of the model specification and it seems that is correct as long as lagged values of the dependent variable were not significant. Results presented in Table 3 indicate that there exists weak evidence in the data for abnormal positive returns at Fridays.

Model (4) was also estimated by robust method with weights and robust standard errors. In this model special attention is paid to observations with small errors, and those with large ones are omitted. In that model 2[nd] lag of dependent variable turns out to be not significant, also the first lag is on the edge of 10% significance level. It seems that on the market exist negative Tuesday effect (at 10% level) and positive Thursday and Friday (both at 2.5%). This suggests that polish stock market in the analysed period was not fully efficient and our result confirm the ones obtained in the literature [Lyroudi et al, (2002), Agrawal and Tandon, (1994)], especially when we consider positive and significant Friday effect.

Model on 5-minute data

Daily effects

In the main scope of the research are HF data based analyses. We divided our HF analysis on daily and hourly effects, additionally presenting results for two separate periods. In the Table 4 estimations of equation (4) on five-minute data are reported. We omit coefficient values for six lags of dependent variable. These lags were included to alleviate eventual autocorrelation problem. In addition, the analysis for each yearly period, starting from June to May the following year were performed. Starting point for the analysis is data driven.

Table 4. Daily effects based on five-minute data models for all analyzed period and two separate periods.

Variable	Standard		Robust		Robust period 0		Robust period 1	
Monday	.0017307		.0022584	***	.0036817	***	.0008559	
Tuesday	-.0004684		.0002019		.0012556	*	-.0008087	
Wednesday	-0005551		-.0004234		-.0014654	**	.0006545	
Thursday	-.0016883		-.0006338		.0006225		-.0018425	**
Friday	-.0003390		.0008401		.0028375	***	-.0001823	
Open Monday	.0319049	***	.0436897	**	.0794787	***	.0030769	
Open Tuesday	1006387	***	.1151625	***	.1594882	***	.0797756	**
Open Wednesday	.0338937		.0650980	***	.108494	***	.0195413	
Open Thursday	.1799356	***	.1831279	***	.0994264	***	.2645347	***
Open Friday	.0814471	***	.0927011	***	.1307886	***	.066525	***

*** 1%; ** 5%; * 10%. Significance level
Source: own computation based on WSE data

First of all, we have to reveal that until July 2007 there was a significant upward trend on WSE and other emerging and developed stock market, which were finished by US subprime mortgage crisis spreading through financial institutions all other the world. Secondly, as one can easily notice from Table 5 there exists persistent open effect on Mondays, strong and positive open effect on Thursdays, and weakling, but positive effect on Fridays. Thirdly, considering the day-of-the-week effect we can not distinguish any effect which is persistence through consecutive years of our research.

Our last observation is in contradiction to the results on daily data where we identified positive and significant Friday effect. On the other hand, we can simply explain this phenomenon stating that the open jump effect is responsible for all day effect.

Table 5. Robust daily effects based on five-minute data models - model with weights and robust standard errors.

Variable	01.06.2003 31.05.2004		01.06.2004 31.05.2005		01.06.2005 31.05.2006		01.06.2006 31.05.2007		01.06.2007 31.03.2008	
Monday	.0056	***	.0003		.0059	***	.0012		-.0018	
Tuesday	-.0017		.0017	*	.0028		-.0008		.0006	

Wednesday	-.0031 *	-.0007	.0028 **	-.0003	-.0014
Thursday	.0013	-.0009	.0025 **	-.0014	-.0058 ***
Friday	.0038 ***	.0022 **	.0023 **	-.0012	-.0009
Open Monday	.1146 ***	.0655 ***	.0953 ***	.2027 ***	-.2627 ***
Open Tuesday	.2768 ***	.0828 ***	.0925 ***	.0575	.1306 ***
Open Wednesday	.1213 ***	.0899 ***	-.0397	.1828 ***	-.0682
Open Thursday	.0825 **	.0992 ***	.1728 ***	.3146 ***	.2316 ***
Open Friday	.1845 ***	.0628 ***	.1515 ***	.0805 ***	.0058

*** 1%; ** 5%; * 10%. Significance level

Source: own computation based on WSE data

Hour effects

Table 6. Intra-daily effects based on five-minute data models for all analyzed period and two separate periods.

Variable	Standard	Robust	Robust period 0	Robust period 1
Open Monday	.0332887 ***	.0431633 ***	.0796399 ***	-.0002165
Open Tuesday	.0999710 ***	.1126434 ***	.1580862 ***	.0761237 ***
Open Wednesday	.0329045 ***	.0624706 ***	.1035983 ***	.0194294
Open Thursday	.1775156 ***	.1804894 ***	.0968423 ***	.2615018 ***
Open Friday	.1810202 ***	.0910487 ***	.1299158 ***	.0635320 ***
09.00-09.30	.0005691	.0025103 ***	.0035837 ***	.0015835
09.30-10.00	-.0034541	-.0023248 ***	-.0019931 *	-.0033376 **
10.00-10.30	-.0020015	.0005255	.0006654	.0005543
10.30-11.00	-.0000939	.0004464	.0000624	.0009186
11.00-11.30	-.0006329	.0002647	.0035878 ***	-.0028283 **
11.30-12.00	.0016751	.0009119	.0012987	.0006247
12.00-12.30	-.0031015 *	-.0020990 ***	-.0022407 *	-.0021691 *
12.30-13.00	.0031205 *	.0028138 ***	.0049659 ***	.0008446
13.00-13.30	-.0009271	.0004872	.0037989 ***	-.0028142 **
13.30-14.00	.0004559	.0004078	.0015297	-.0008244
14.00-14.30	-.0009303	.0006573	.0007367	.0006148
14.30-15.00	-.0014699	-.0008185	.0005208	-.0021163
15.00-15.30	.0002428	.0021025 **	.0021027 *	.0021978 *
15.30-16.00	-.0014326	-.0017034	-.0007450	-.0027457 *
16.00-16.30	.0169178 ***	.0152226 ***	.0162160 ***	.0150186 ***

*** 1%; ** 5%; * 10%. Significance level

Source: own computation based on WSE data

Open jump is significant and positive on all days but Monday and Wednesday in the second period (Period 1). Moreover, the end of session effect is significant and positive, despite period analysed or chosen method of estimation. The significant value for 15:00-15:30 and 15:30-16:00 intervals may be linked with an effect of the anticipation of and reaction for NYSE opening but these effects are not persistent when we conduct our analysis separately on two time periods.

Mid–session significant results are only statistically different from zero for the first period, i.e. for the period when the future market was young and emerging. The existence of these effects we can explain on the basis of Macro data announcement which were revealed by the Main Statistical Office at many different hours of the day (10:00, 11:00, 13:00, !4:00, etc).

Table 7. Intra-daily effects based on five-minute data models (Newey-West regressions)

Variable	Robust	Robust period 0	Robust period 1
Open Jump	.0899028 ***	.1075805 ***	.0719269 *
09.30-10.00	-.0068127		-.0036315
10.00-10.30	-.0017638	-.003567	-.0032355
10.30-11.00	-.0002819	-.0022015	.0015484
11.00-11.30	-.0003506	.0044112 **	-.0048052 **
11.30-12.00	.0013597	.001302	.0016386
12.00-12.30	-.0033257 **	-.0040781 **	-.0023849

12.30-13.00	.0030934	**	.0051013	***	.0008165	
13.00-13.30	-.0008091		.0025709		-.0040103	**
13.30-14.00	.0004011		.0004852		.0001893	
14.00-14.30	-.0008872		-.0023798		.0004827	
14.30-15.00	-.0016025		-.0015225		-.001535	
15.00-15.30	.0002379		-.0007748		.00113	
15.30-16.00	-.0014896		-.001043		-.0025331	
16.00-16.30	.0171273	***	.021536	**	.0158256	**

*** 1%; ** 5%; * 10%. Significance level
Source: own computation based on WSE data

When autocorrelation is controlled by Newey-West standard errors with 36 lags, beside open jump and end of session effect, two returns on mid-session periods are statistically significant but not persistent. Therefore, after controlling of possible autocorrelation of order 36 we found similar results to the previous ones.

When we take a closer look at results of robust estimation of daily effects we can easily discover that there is a shift from day effects to open effects. As a market grows it is supposed to become more similar to his developed counterparts, revealing their characteristics as well. However, strong open effects on all days but Wednesday, and at the end of session effect remain persistent.

Table 8. Intra-daily effects based on five-minute data models (Newey-West regression)

Variable	Standard		Robust	
Open Monday	.0658942	***	.0807635	***
Open Tuesday	.0591933	***	.0664849	***
Open Wednesday	-.011504		-.0025753	
Open Thursday	.1760493	***	.1809812	***
Open Friday	.1604134	***	.1734226	***
09.30-10.00	-.0065774	**	-.0010692	
10.00-10.30	-.0015168		.0012462	
10.30-11.00	-.0003997		.0005782	
11.00-11.30	-.0004671		.0004218	
11.30-12.00	.0012664		.0009079	
12.00-12.30	-.0032964		-.0022638	***
12.30-13.00	.0031173		.0027322	***
13.00-13.30	-.0008653		.0006154	
13.30-14.00	.000392		.0004083	
14.00-14.30	-.0008671		.0007486	
14.30-15.00	-.0015678		-.0009256	
15.00-15.30	.0002608		.0021406	**
15.30-16.00	-.0013289		-.0016553	
16.00-16.30	.0167264	***	.0150899	***

*** 1%; ** 5%; * 10%. Significance level
Source: own computation based on WSE data

Our last Table 9 presenting results of intra-daily analysis informs us that intra-day mid-session effects are not stable when we analyse subsets of data. It is hard to judge if this instability comes from a smaller database or differences between periods (i.e. strong and significant upward trend in the first phase of our research or strong downward movement in the second phase). On the other hand, next time we observe significant and positive end of session effect and open effect, which is persistent for Monday, Thursday and Friday.

Table 9. Intra-daily robust effects based on five-minute data models

Variable	01.06.2003-31.05.2004		01.06.2004-31.05.2005		01.06.2005-31.05.2006		01.06.2006-31.05.2007		01.06.2007-31.03.2008	
Open Monday	.2359	***	.0623	**	.1850	***	.1519	***	-.2629	***
Open Tuesday	.1684	***	.0040		.0570	**	.0052		.0616	
Open Wednesday	.0326		-.0210		-.1453	***	.1381	***	-.1456	***

Open Thursday	.0441		.1692	***	.1357	***	.3570	***	.2419	***
Open Friday	.3345	***	.1617	***	.1338	***	.1677	***	.0513	*
09.30-10.00	-.0002		.0013		.0059	*	-.0018		-.0023	
10.00-10.30	.0001		.0021		.0028		-.0035	*	.0025	
10.30-11.00	-.0013		-.0006		.0015		.0034	*	-.0009	
11.00-11.30	.0082	***	-.0001		.0033	*	.0002		-.0086	***
11.30-12.00	.0013		.0009		.0006		.0023		-.0008	
12.00-12.30	-.0056	***	-.0014		-.0002		.0011		-.0063	***
12.30-13.00	.0026		.0050	***	.0065	***	-.0006		-.0015	
13.00-13.30	.0047	**	.0040	**	.0027		-.0027		-.0079	***
13.30-14.00	.0034	*	-.0029	*	.0063	***	-.0030		-.0025	
14.00-14.30	-.0012		.0002		.0005		.0022		.0025	
14.30-15.00	.0016		-.0013		.0005		-.0050	**	.0002	
15.00-15.30	-.0010		.0050	***	.0047	**	.0016		-.0011	
15.30-16.00	-.0003		-.0023		.0029		-.0032		-.0080	***
16.00-16.30	.0378	***	.0058		.0087	**	.0121	***	.0236	***

*** 1%; ** 5%; * 10%. Significance level

Source: own computation based on WSE data

5. Summary and questions for future research

Taking into account all the results presented above we can formulate the following conclusions:. Firstly, there exists strong hour of the day effect revealed in two effects connected with the start and end of the stock session:

- the open jump effect, which is persistent for Monday, Thursday and Friday;
- the end of session effect, which is additionally persistent.

Secondly, the results confirmed our initial supposition about significant open jump and end of day effect closely connected with substantial fluctuations of volatility at the same time (Figure 3.2).

Our next conclusion concerns the shift from day-effect into open jump effect, i.e. the day-of-the-week effect is restricted only to open jump effect or on the other hand the uncertainty between the close and open on the following day is revealed just after the market open. Such an explanation presenting the example of market inefficiency, on the other hand, informs us about much faster incorporation of the new information into the traded prices in comparison to the daily effect.

When we focus on the methodology applied in our research we can additionally notice that the results are resistant, because we eliminated an influence of potential outlying observations which usually dominate the data. Additionally, we employed robust estimators for variance-covariance matrix in order to obtain more reliable estimators of variance what enabled us to reveal presented effects more precisely.

Revealed patterns do not necessary be persistent. We are aware of the fact that inefficiencies found in the data will diminish while the market become less emerging market and more developed but we have to stress that it is still not the case of WSE.

At the end we would like to describe some paths for future research into this subject which could enable finding the answer for the degree of emerging market efficiency and the stability of our results. In the future researches more emphasis should be put to the analysis of higher frequencies (tick data). Such analysis could identify the causes of open and final jumps, additionally, explaining inefficiency on the ground of market microstructure. Finally, more evidence concerning the intra-day effects should be revealed while analysing the volatility on the base of high frequency data what will be the subject of our next research.

References

[1] Agrawal A., Tandon K., (1994), *Anomalies or Illusions?: Evidence from Stock Markets in Eighteen Countries*, in: *Journal of International Money and Finance*, pp. 83 – 106.

[2] Ariel R., (1987), *A Monthly Effect on Stock Returns*, in: *Journal of Financial Economics*, Vol. 18, pp. 161 – 174.

[3] Bachelier L., (1900), *Theory of Speculation*, Gauthier-Villars, Paris, in: *The random character of stock market prices*, Cootner P., MIT Press, Cambridge Mass., (1964), pp. 17 – 78.

[4] Bhattacharya K., Sarkar N., Mukhopadhyay, (2003), *Stability of the day of the week effect in return and in volatility at the Indian capital market: an ARCH approach with proper mean specification*, in: *Applied Financial Economics*, Vol. 13, pp. 553 – 563.

[5] Buczek S., (2005), *Efektywnosc informacyjna rynkow akcji – teoria a rzeczywistosc*, in: *Oficyna Wydawnicza SGH*, Warszawa.

[6] Campbell J.Y., Lo A.W., MacKinlay A.C., (1997), *The econometrics of Financial Markets*, Princeton University Press.

[7] Cross F., (1973), *The Behavior of Stock Prices on Fridays and Mondays*, in: *Financial Analysts Journal*, Nov-Dec., pp. 67 – 69.

[8] Czekaj J., Wos M., Zarnowski J., (2001), *Efektywnosc gieldowego rynku akcji w Polsce*, PWN, Warszawa.

[9] Das S., Arora V., (2005), *Day of the week effects in NSE stock returns*, SSRN working paper.

[10] Fama E., (1965), *The Behaviour of Stock Market Prices*, in: *Journal of Business*, Vol. 38, pp. 134 – 105.

[11] Fama E., (1965), *Random Walks in Stock Market Prices*, in: *Financial Analysts Journal*, September/October.

[12] Fama E.F., (1970), *Efficient capital markets: A review of theory and empirical work*, in: *Journal of Finance*, Vol. 25, No. 2, pp. 381 – 417.

[13] Fama E.F., French K.R., (1988), *Permanent and Temporary Components of Stock Prices*, in: *Journal of Political Economy*, Vol. 96, pp. 246 – 273.

[14] Guin, L, (2005), *Handout on market anomalies in the course Investment Management*, Professor in finance, Department of Economics and Finance, Murray State University.

[15] Harris L., (1986), *A Transaction Data Study of Weekly and Intradaily Patterns in Stock Returns*, in: *Journal of Financial Economics*, Vol. 16, pp. 99 – 117.

[16] Haugen R., Lakonishok J., (1988), *The incredible January effect*, Dow Jones-Irwin, Homewood.

[17] Harvey, C.R., Huang, R.D., (1991), *Volatility in the foreign currency futures market*, in: *Review of Financial Studies*, Vol. 4, No. 3, pp. 543 – 569.

[18] Jajuga K., (2000), *Metody ekonometryczne i statystyczne w analizie rynku kapitałowego*, Wydawnictwo AE we Wroclawiu, Wroclaw.

[19] Jegadeesh N., Titman S., (1993), *Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency*, in: *The Journal of Finance*, Vol. 48, No. 1, pp. 65 – 91.

[20] Keim D., Stambaugh, R., (1984), *A Further Investigation of the Weekend Effect in Stock Returns*, in: *Journal of Finance*, July, pp. 819 – 837.

[21] Koenker R., Portnoy S., (1987), *L-estimation for linear model*, in: *Journal of American Statistical Association*, Vol. 82, pp. 851 – 857.

[22] Lakonishok J., Levi M., (1982), *Weekend effect on stock returns. A note*, in: *The Journal of Finance*, Vol. 37, No. 3, pp. 883 – 889.

[23] Lakonishok J., Smidt S., (1988), *Are Seasonal Anomalies Real?: A Ninety Year Perspective*, in: Review of Financial Studies, Vol. 1, pp. 435 – 55.

[24] Lakonishok J., Maberly E., (1990), *The weekend effect: trading patterns of individual and institutional investors*, in: Journal of Finance, Vol. 45, No. 1, pp. 231 – 243.

[25] Lehman B.N., (1990), *Fads Martingales and Market Efficien*cy, in: The Quarterly Journal of Economics, Vol. 105, No. 1, pp. 1 – 28.

[26] Lyroudi K., Subeniotis D., Komisopoulos G., (2002), *Market Anomalies in the A.S.E: The Day of the Week Effect*. EFMA 2002 London Meetings, February 25.

[27] Murinde V., Pashakwale S., (2002), *Volatility in the Emerging Stock Markets in Central and Eastern Europe: Evidence on Croatia, in: Czech Republic, Hungary, Poland, Russia and Slovakia*, European Research Studies Journal.

[28] Nath G.C., Dalvi M., (2004), *Day of the week effect and market efficiency-evidence from Indian equity market using high frequency data of national stock exchange*, working paper.

[29] Pashakwale S., Murinde V., (2001), *Modelling the Volatility in East European Emerging Stock Markets: Evidence on Hungary and Poland*, in: *Applied Financial Economics*, 11, pp. 445 – 456.

[30] Poterba J.M., Summers L.H., (1988), *Mean Reversion in Stock Prices: Evidences and Implications*, in: *Journal of Financial Econmics*, Vol. 22, No. 1, pp. 27 – 59.

[31] Rozeff M.S., W.R. Kinney, (1976), *Capital Market Seasonality: the Case of Stock Returns*, in: *Journal of Financial Economics*, Vol. 3, pp. 379 – 402.

[32] Samuelson P., (1965), *Proof that properly anticipated prices fluctuate randomly*, in: *Industrial Management Review*, Vol. 6, pp. 41 – 49.

[33] Ślepaczuk R., (2006), *Technical Trading Strategies and Market Efficiency*, in S. Motamen-Samadian: *Global Stock Market and Portfolio Management*, Palgrave Macmillan, New York,

[34] Slepaczuk R., Zakrzewski G., (2008), *The daily volatility patterns of realized volatility*, manuscript in progress.

[35] Smirlock M., Starks L., (1986), *Day-of-the-week and intraday effects in stock returns*, in: *Journal of Financial Economics*, Vol. 15, pp. 296 – 421.

[36] Szyszka A., (2003), *Efektywnosc GPW w Warszawie na tle rynkow dojrzalych*, AE w Poznaniu, Poznan.

[37] Wood R., McInish T., Ord K., (1985), *An investigation of transaction data of NYSE stocks*, in: *Journal of Finance*, Vol. 40, No. 3, pp. 723 – 739.

Appendix

Figure 3: Daily log-returns

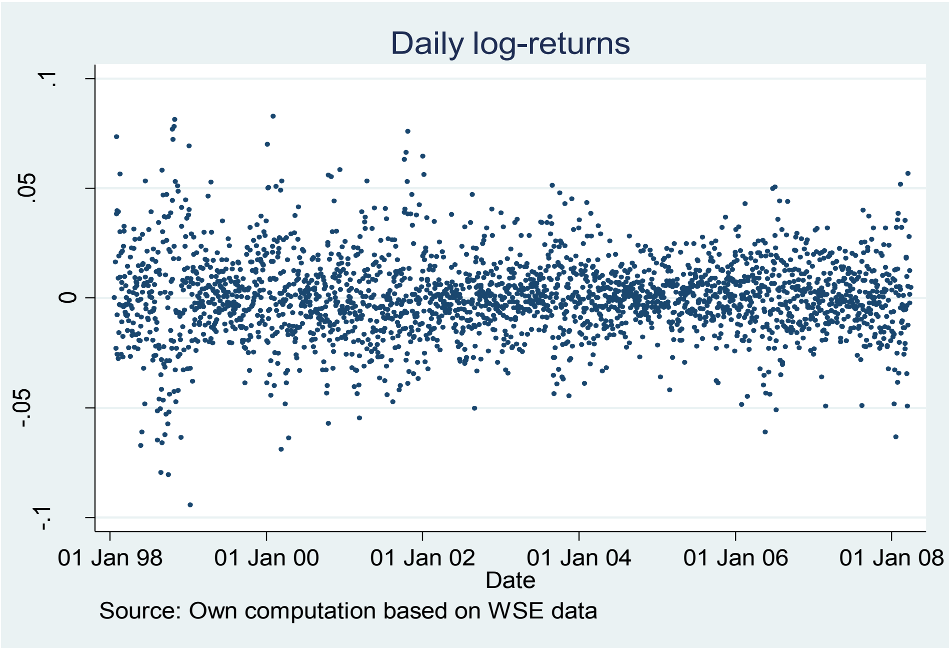


Figure 4: Daily log-returns squared

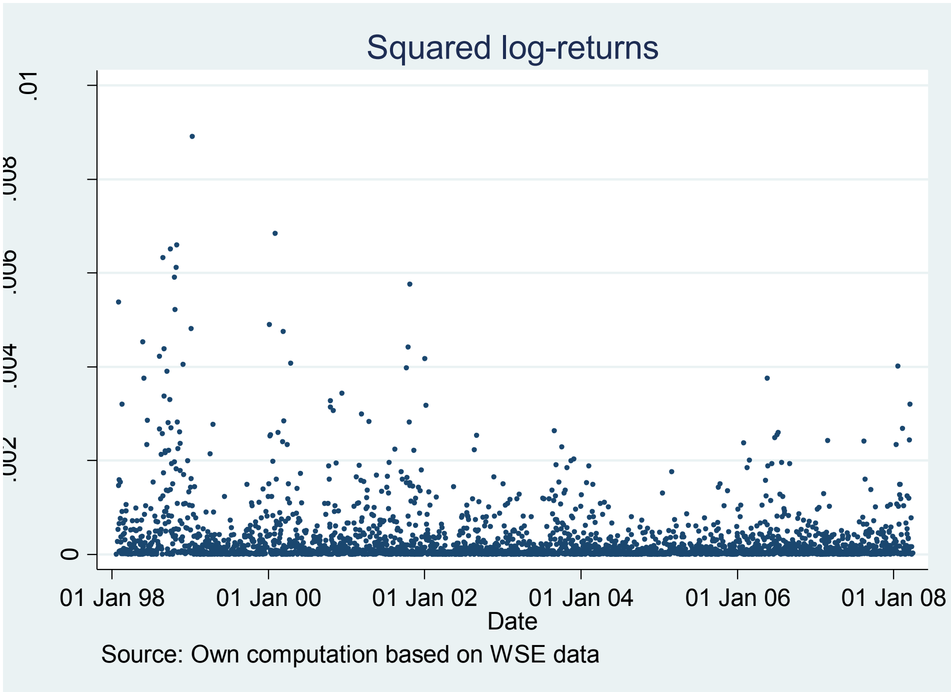
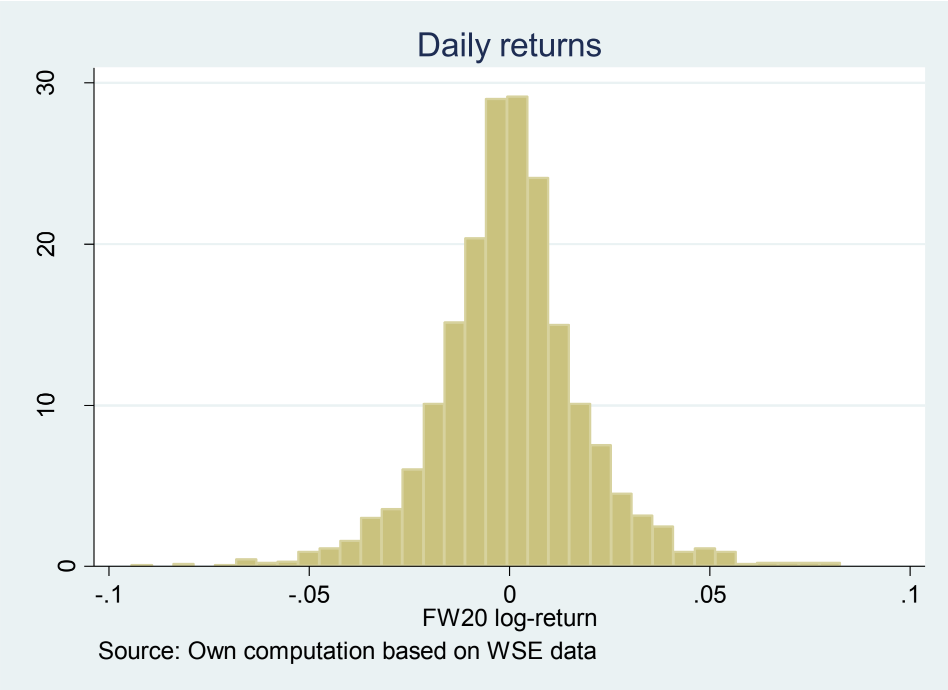


Figure 5: Histogram of daily log-returns



A SIMPLE MODEL OF DECISION MAKING – HOW TO AVOID LARGE ERRORS?

Zoltan **VARSANY**
Research Department, Standard and Poor's
varsanyiz77@yahoo.com

Abstract:

In this paper I present a simple model through which I examine how large unwanted outcomes in a process subject to one's decisions can be avoided. The paper has implications for decision makers in the field of economics, financial markets and also everyday life. Probably the most interesting conclusion is that, in certain problems, in order to avoid large unwanted outcomes one, regularly and intentionally, has to make decisions that are not optimal according to his/her existing preference. The reason for it is that the decision rule might get "overfitted" to one's (recent) experience and may give wrong signals if there is a change, even as temporary as in one single period, in the environment in which decisions are made. I find the optimal decision making strategy in an example case – the optimal strategy, however, may well be different in different real-world situations.

Keywords: endogeneity, non-stacionarity, outliers, simulation, uncertainty

JEL Classification: C15, D81

1. Introduction

In this paper I examine, through a very simple model, how large outliers in a process may emerge and can be avoided by a decision maker. On the one hand, modelling in the financial markets became increasingly challenging (although I think the modelling of a single interest rate can be quite challenging in itself), especially by the introduction of more and more complex instruments, which may question the validity of the present models. One does not have to go far into the past: models for such complex instruments as mortgage backed securities and CDO-s, for example, are partly blamed for the recent turmoil on credit markets related to US mortgages (see, for example, out of many similar commentaries, Greenspan (2008)).

Still, the models are used and, especially, decisions are based on their results. This latter point was emphasised in a recent book, Rebonato (2007), which was part of the inspiration behind the model presented here (and why the emphasis here is not on forecasting but on decision making). For some time I have found it an interesting question what the effect of decisions based on incorrect models can have. More specifically, what repercussion effect such decisions can have on the system as a whole. Is the application of these decision rules (which are, in essence, the decision makers' models of the process in question) going to form the process in such a way that the rules become "true" even if they were not at the beginning? For example, is technical analysis of share prices self-fulfilling? Or, how can one judge the effectiveness of the decisions about the interest rates of central banks?

Two other related problems are that of building such a decision rule and the role past data should play in it. Taleb (2004) strongly reinforces one's scepticism of some models based on past data. This appears in two aspects in this paper: first, it turns out to be an important factor in the model presented here how far the decision maker looks back in the history of the data. More important is, however, the question how we could avoid unexpected outcomes of large magnitude without knowing too much about the "external" dynamics of the process (i.e. the dynamics that would prevail if there were no external action to it). A difference to Taleb's discussion is that while there the outliers are given externally (and the decision maker can only act passively to avoid large negative outcomes), here they are endogenous (but the decision maker doesn't necessarily know in advance when they are going to appear).

This last point leads us to what the paper is about and what it isn't. The primary reading of this paper is that it's about a decision maker who *can influence* (actually, is the only actor to have an influence on) a given process and tries to set it to a level determined as a target. He/she thus is best viewed as a macro-economic decision maker. To some extent the model can also be viewed as describing the dynamics of prices on financial markets – although I can't give a perfect correspondence between the model's variables and relationships and those of real markets. The decision maker might be viewed as the market as a whole (the investors), the target as the market "sentiment" or expectation and the decision rule as the market's past experience as to the effects of news on market prices. This latter factor would then work as follows: given past news and the subsequent changes in prices the market can estimate how future news may affect prices. If, over a long period, no important news is coming the market "forgets" how itself is reacting to "big" news, which is a potential source of overreaction.

It is important to see what the paper does *not* address: it does not try to give a model that would give the best possible description and – however dubious – prediction of a given market or variable. In this respect it is thus different from some part of the existing literature in which complicated stochastic processes, many variables, utility functions, etc. are at work. An example, from a field which is quite concerned about "outliers" or large sudden changes, is Sornette (2003), where the author develops a model of financial market bubbles – and the reader might have the impression that it is more important that the model mimics past asset price patterns than it's actual predictive power (and, sometimes, intuition).

Indeed, this paper focuses on decision making rather than on prediction: we should find such decision rules by which we can either avoid large unwanted outcomes (as in the model presented here), or withstand those outcomes (as an option buying trader in Taleb (2004)).

The purpose of this paper is to show that even the simplest models can be useful in getting insight into more complicated decision making problems. The paper doesn't try to develop a decision making theory, not even a whole class of models. One simple structure is presented here, noting that this, itself, can lead to very different dynamics and outcomes, depending on how one formulates the building blocks of the structure.

2. The model

First we have the "external" dynamics of the system:

$$x_t = \alpha(x_{t-1} + a_t) + e_t \tag{1}$$

with e distributed normally. Then, we have the DM's decision rule:

$$a_t = (E_t - x_{t-1} - \beta)/\gamma \tag{2}$$

where β and γ are the intercept and slope, respectively, of the regression of the past changes of x on the past actions of the same period (a),

$$x_t = \beta + \gamma a_t + v_t \tag{3}$$

E_t is the DM's expectation/target of x in period t set in period t-1. Thus, the DM uses his past experience about the results of his actions to find his current period action such that this action will drive the process to the expected/required level.

Apart from the above few parameters we have only a few additional factors to set. First, the number of periods the regression determining the DM's action, (3), looks back. I will examine two possibilities: "long memory", looking back till the beginning of the sample period and "short memory", looking back for 5 periods. Another parameter is found in the determination of the expectation:

$$E_t = E_{t-1} + \delta(x_{t-1} - E_{t-1}) \tag{4}$$

thus δ represents the adaptation of the expectation to the actual value of the process. Importantly, if δ is set to zero, the model represents a situation where the DM has a fixed target for the value of x.

I also examine an alternative form of the evolution of the expectations:

$$E_t = E_{t-1} + \varepsilon_t \tag{4.1}$$

i.e., E is a random walk (ε is distributed normally).

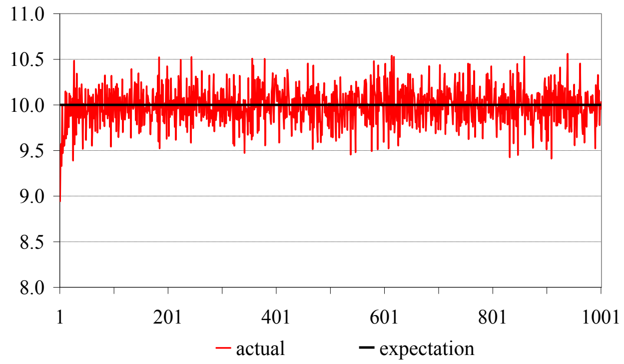
The model runs as follows. At the end of a given period (let's say, period t) the DM observes the outcome of the process, x_t, and how far away it is from its expectation or target (E_t). It then calculates its decision rule based on Eq. (3), and using this rule it calculates its required action for the next period (a_{t+1}) that is supposed to move the process to its expected/target level (E_{t+1}, determined by (4) or (4.1)). Finally, in period $t+1$, the DM's action and the "external dynamics" of the system together bring about the new value of the process, x_{t+1} – and the circle starts again.

3. Results

In this part I show the results of runs with different set-ups. There will be one common part in each scenario, however: the "external" dynamics, in (1), will have α = 1.2, that is, without the DM's action the process explodes. This part is just to give a "feeling" of how the model works. The most interesting conclusion is going to be drawn in the next section.

With a fixed target (δ = 0) and long memory the process nicely fluctuates around the target value, set at 10 in the below example:

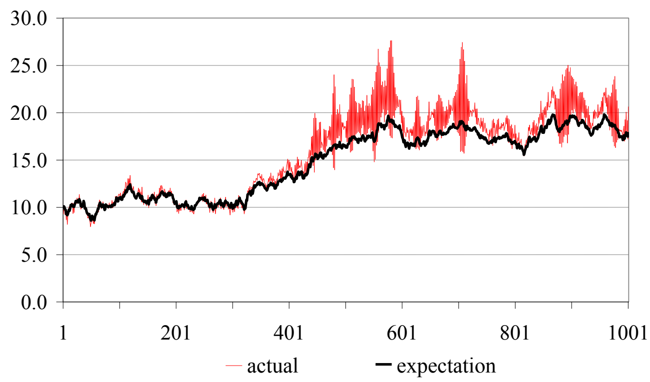
Figure 1: The actual value of the process with fixed expectation (which equals 10) and long memory



Introducing a moving target (δ > 0) will result in the process exploding. This is because the DM's actions are going to be smaller in absolute size then necessary to keep the process within constraints.

Finally, if the expectation is evolving according to a random walk, with long memory we can see that it has a remarkable effect on the process:

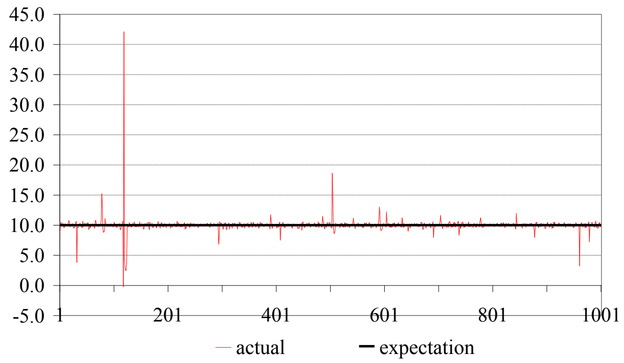
Figure 2: The actual value of the process with random walk expectation and long memory



Interestingly, we can see nice volatility clustering in the change of the actual process – and the process is still contained.

With short memory, we get qualitatively quite different results. With fixed target the process will, in general, be much closer to the target than with long memory. However, there will be "devastating" outliers:

Figure 3: The actual value of the process with fixed expectation (which equals 10) and short memory



Though these outliers will still be there, the process won't explode even with flexible targets:

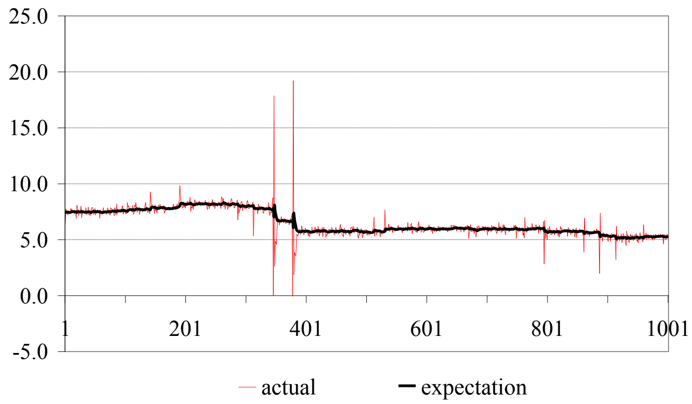


Figure 4. The actual value of the process with adaptive expectation and short memory

Finally, the same happens with random walk expectations:

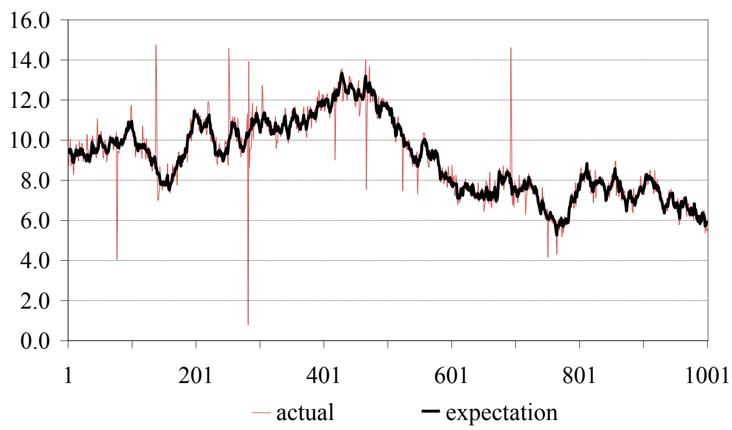


Figure 5. The actual value of the process with random walk expectation and short memory

To conclude this section we can have the following important observation. Apart from the adaptive-expectation case both the long and the short memory decision maker can contain the process (it doesn't explode). However, there is a qualitative difference: with long memory the errors made by the DM seem to be more homogenous while in the short memory case there can be huge outliers. At the same time, apart from the outliers, short memory seems to lead to higher accuracy (a smaller average distance between the target and the outcome).

In what follows I will examine the short memory cases. The reason for it is that, on the one hand, it produces such outliers that we observe in many real-life areas. On the other hand, short memory doesn't seem to be a very unrealistic assumption. It models, for example, "fast changing world" (changing environment in which decisions have to be made), changes in the person of the decision maker (the new DM has less experience), and so on.

4. How to avoid the large outliers?

In this section I'm examining the major question of the paper. In the previous section we observed that with short memory we can have large outliers in the difference between the expectation and the actual outcome caused by the DM's inappropriate action. At the same time we had the impression that the short memory leads to smaller differences on average. Consequently, based on some cost function a particular decision maker might want to choose to have short memory – now let's assume that it is not possible: above a certain level of the (absolute) difference between the target and the outcome the DM "explodes" (or is replaced by a more competent one…). Furthermore, let's assume that long memory is not an option any more. So what can the DM with short memory do to avoid the outliers?

Let's try to understand the difficulty of the problem first. Let's say, we are the decision maker. We know the past decisions and realizations of the process and have a very simple model of their recent relationship (we know the whole past but use only the past few observations). We know nothing else. We don't know the "external" dynamics; we can't have valid forecasts; we don't know when and why outliers came in the past. There is a chance that if we decide based on our model an outlier comes (and from that point on we will have more time fishing). I assume in such a situation the DM would not risk deciding upon its model. So what can he do? No more data, no more knowledge of the "external" dynamics.

For now, let's use fixed expectations! A first good idea is to check why in the past we had outliers. It turns out that estimating the expectation-regression with only 5 data-pairs (let's define short memory in this way) the parameters can have such "strange" values that lead the DM to an incorrect decision. The following figure shows the values of the process in each period (x-axis) against the slope in the preceding period of the regression on which the expectation is based. It is obvious that the outliers are caused by a very small slope that leads to an overreaction by the DM.

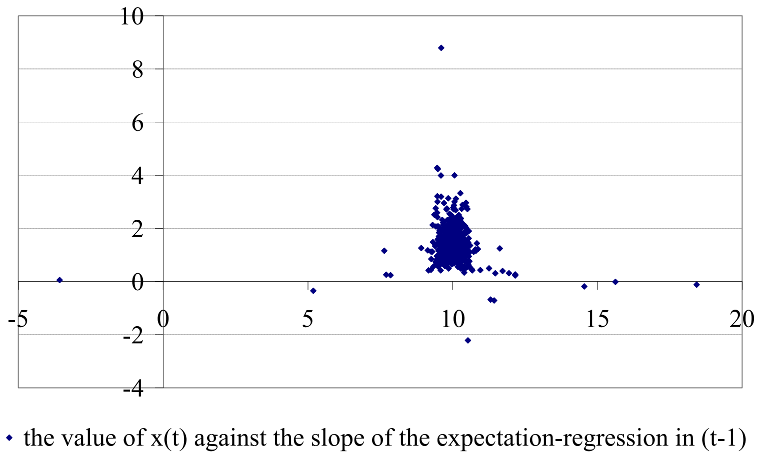


Figure 6. Outliers (x-axis) are caused by a very tiny (in absolute value) slope of the decision equation (y-axis)

But for now, let's assume that the DM doesn't actually know what was wrong – only has some ideas that his model doesn't always give good advice because of the parameters are not always correct (but the DM doesn't know when).

What it can do then is to change it's decision scheme. So far it was assumed that the decisions are "homogenous" in the sense that in each period the same rules were applied: take the past few observations; calculate the regression slope and intercept; use these parameters to calculate how much action is needed to achieve the target.

Now let's use a different rule: in every second period, instead of deciding based on the regression model, let the DM decide *not* based on the model. First I examine the case when the DM applies the same fixed action in every second period: the values of -4, -3, -2, -1 and 0 are considered. It turns out that starting with -4 first the outliers are eliminated and the errors decrease in absolute value; then approaching, most probably, -1.66 larger and larger outliers appear again; finally increasing towards zero (and above) outliers die out and the average error increases. The role of -1.66 is quite possibly that it is the "steady state" solution of the "external" dynamics if the fixed target is 10.[42] The following figures demonstrate these findings:

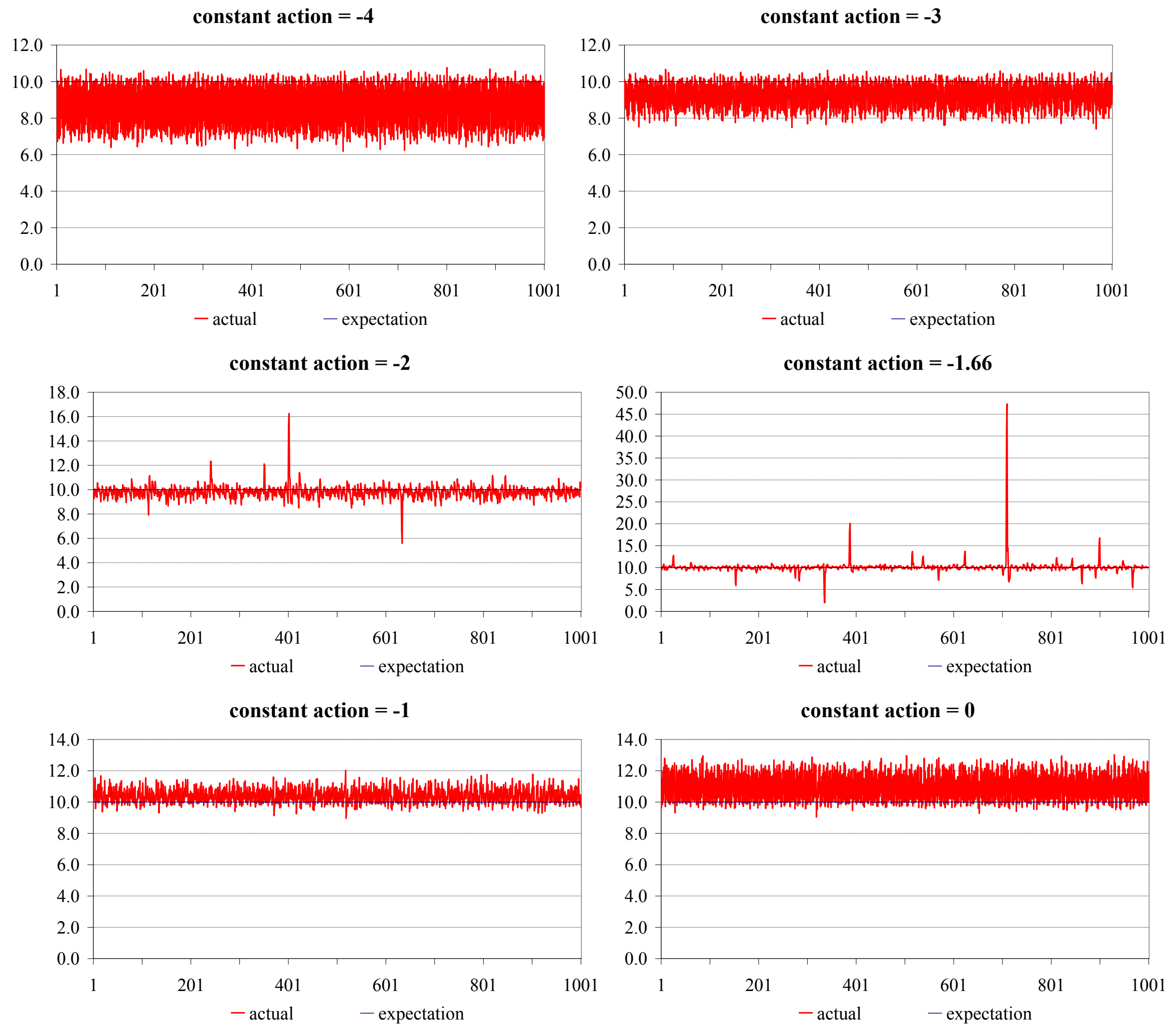


Figure 7. Errors accompanying different model-independent decision levels within the modified decision rule

What are the implications? A decision maker with a fixed target and a decision rule which is based on its past actions and the corresponding outcomes can either choose relative accuracy most of

[42] From (1): $10 = 1.2(10 - a_t)$ it follows that $a_t = 5/3$.

the time and high errors in some periods or small errors in every period. The steady, small errors are a result of its way of making decisions: it uses not only its model (the regression) but, in given periods, it uses also another, model independent decision rule (a constant) such that the actions based on this latter rule and the corresponding outcomes help to form the original model correctly.

An important question now is how the DM can determine the "appropriate" value for the constant to be applied. Remember: the DM wants to stay alive – this may not be very easy at the beginning when it can hardly have any idea whether it should apply, say, 100 or -20 to get a reasonable outcome. Since a decision *is* made in each period (even zero action is a decision), there can't be a guarantee that a fully inexperienced DM will survive the first couple of periods – I think no such decision rule can be found. However, if the DM has at least *some* experience, even as few as 2 or 3 periods, it can have some ideas as to what levels of action can be applied with the probability of exploding being low. Most importantly, the initial actions should not be "too" close to each other otherwise the extrapolation based on these to action/outcome pairs may turn out to be very misleading. In fact, simulations suggest that given the DM survives the first two decisions, as little as this idea can ensure it stays alive.

With flexible targets (or expectations) the method we found in the fixed target case doesn't work any more. With adaptive expectations – using (4) – applying a fixed action in every second period will make the process explode. This is because as the absolute value of the process increases, the absolute value of the change of the constant action also increases. This will cause the decision regression to move up/down steadily (i.e. its intercept increases in absolute value). This is demonstrated in the figure below:

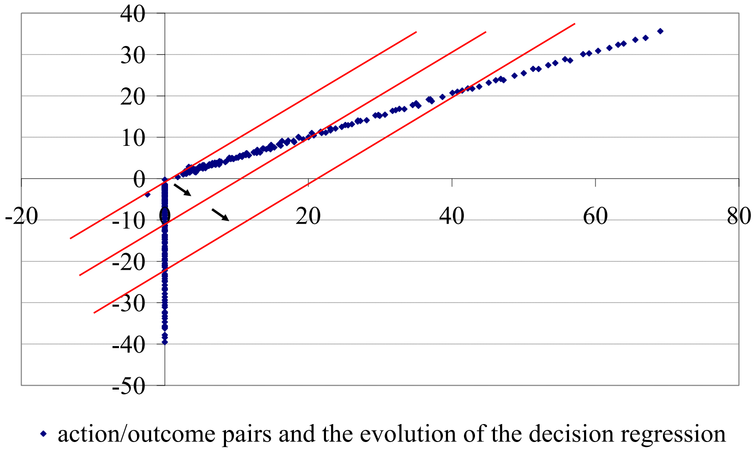


Figure 8. The way the process explodes with short memory, adaptive expectations and the modified decision rule

After each model independent action the constant increases in absolute value but the regression for the next period can't adapt to it perfectly – as a result the action will be smaller than necessary. This, in turn, will result in a further up/down adaptation of the expectation and everything starts again.

As it turns out, in the flexible target case, it is much more difficult to avoid the outliers in the process. The DM even runs the risk that the process explodes – however, this can be avoided by a new simple rule: the DM has to act in every second period exactly as it acted in the immediately preceding period. This will result in outcomes such as the one in the figure below:

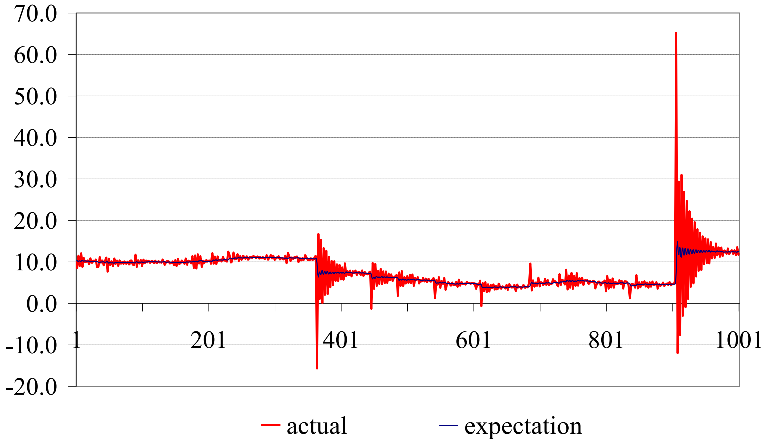


Figure 9. A further modification of the already modified decision rule helps to contain the process but outliers can not be avoided.

As we see, the outliers are still there (however, the process is contained, at least).

Finally, with flexible targets and expectations that evolve according to a random walk the constant decision rule that was successful in the fixed target case works. As the constant decision approaches -1.66 from below the approximation of the target becomes better in most of the cases but the size of the outliers also increases; increasing beyond -1.66 outliers gradually vanish and the average approximation error increases.

5. Conclusion

Above, I modeled a situation when a decision maker has relatively few information about the process it wants to affect and still wants to avoid large unexpected outcomes. The model was very simple on the one hand but quite specific on the other: it used a non-stacionary regression as the "external dynamics" of the process. Thus, my conclusions drawn above are possibly constrained to be valid only for this specific model. Probably, if I had used a higher order polynomial for the external dynamics, for example, I would have gotten very different optimal rules.

However, it is worth pointing out two things. First, even this simple model can result in very different behaviors, that is, processes with completely different time-series properties. This is because although the model's dynamics is driven by an essentially one-parameter equation (the variance of the error term is kept constant) we have some other factors to drive the results: the memory of the DM, its decision rule and its target.

Second, and more importantly, the results do give us some general hints as to the treatment of the risk of large unwanted outcomes. I think it's a quite general situation when one has to decide based on his/her own experience (although, of course, it's not always the case). Then, the model tells us that when considering a decision we should not only think of the most optimal outcome of the current decision but also of how this decision/outcome pair can improve our decision rule in general. With other words, it might be optimal in some periods to base the decision not on the model of previous actions and outcomes but on an external model or rule which is independent of the "main" decision rule.[43]

In this particular model the problem stems from the (small) error term in the process: when the DM's target and the process are close to each other the parameters in the decision regression are mostly influenced by noise. Then, it can easily happen that different (small) actions go with very similar outcomes, thus signaling to the DM that the system is not sensitive to its decisions at that level. When an error term of a slightly higher magnitude arrives the DM will overreact to it and this will drive the process away from the target and may cause large outliers.

[43] One can think of, for example, regulation of credit risk. The more risk sensitive the regulation, the higher the need for accurate modelling. When the author was a member of a European-level working group on large exposures he took part in debates about how risk-sensitive large exposure rules should be. It is more or less known: the large exposure rules are far less risk sensitive – that is, more model-independent – than the Basel II rules in general.

More generally, the model suggest that it can sometimes be worthwhile to choose an option that is not the best one based on our criteria: we can thus discover new action/outcome pairs which improve our future decisions through more experience.

6. References

[1] Greenspan, A., (2008), *We will never have a perfect model of risk*, in: *Financial Times*, March 16, 2008.

[2] Rebonato, R., (2007), *Plight of the Fortune Tellers: Why We Need to Manage Financial Risk Differently*, Princeton University Press

[3] Sornette, D., (2003), *Critical Market Crashes*, http://arxiv.org/abs/cond-mat/0301543. Accessed on 10 February 2008.

[4] Taleb, N.N., (2004), *Fooled by randomness*. Penguin Books.

A NEW MODEL FOR STOCK PRICE MOVEMENTS

Guido **VENIER**
Financial Risk Controller Department
Dresdner Kleinwort Investmentbank[44], **GERMANY**
guido-venier@freenet.de

Abstract:
This paper presents a new alternative diffusion model for asset price movements. In contrast to the popular approach of Brownian Motion it proposes Deterministic Diffusion for the modelling of stock price movements. These diffusion processes are a new area of physical research and can be created by the chaotic behaviour of rather simple piecewise linear maps, but can also occur in chaotic deterministic systems like the famous Lorenz system. The motivation for the investigation on Deterministic Diffusion processes as suitable model for the behaviour of stock prices is, that their time series can obey mostly observed stylized facts of real world stock market time series. They can show fat tails of empirical log returns in union with timevarying volatility i.e. heteroscedasticity as well as slowly decaying autocorrelations of squared log returns i.e. long range dependence. These phenomena cannot be explained by a geometric Brownian Motion and have been the largest criticism to the lognormal random walk. In this paper it will be shown that Deterministic Diffusion models can obey those empirical observed stylized facts and the implications of these alternative diffusion processes on economic theory with respect to market efficiency and option pricing are discussed.

Keywords: deterministic diffusion, stock pricing, fat tails, heteroscedasticity, long range dependence, option pricing.

JEL Classiffication: G12

1. Introduction

Despite the popularity of normal distributed log returns to model the movement of stock prices there has been a large discussion in literature whether this model is appropriate or not. A large bulk of empirical research has been brought forward that mainly concludes that stock market returns are not normally and independently distributed and hence do not follow random walks. For an example see [1]. It is also a fact that so far no satisfying timeseries-model exists for the concurrent explanation of common stylized facts of stock market time series. That are: 1.) fat tails, 2.) heteroscedasticity and 3.) long range dependence. Even stochastic models cannot always accommodate all stylized facts at a time sufficiently. For an extensive discussion please refer to [2].

2. Deterministic Diffusion

Firstly Deterministic Diffusion will be defined and secondly simple piecewise linear maps will be presented that generate such time series.

Definition 2.1 *(Deterministic Diffusion):*

Deterministic Diffusion is the displacement of a particle X on the real line in time according to a deterministic law:

$$X(t+1) = M\left(X(t)\right) \tag{1}$$

Where X(t) denotes the position of the particle at time t and M is a deterministic mapping. A mapping is called expanding if. M´> 1.

Call $\Im$ the family of piece wise linear maps M:
$\mathbb{R} \rightarrow \mathbb{R}$/ith uniform slope s having the properties:

[44] Any views expressed are solely the views of the author and not those of the firm.

1. M is expanding: $s > 1$.
2. M is lifting: $M(X-n)+n=M(X)$ for any real number X and integer n with $n=\mathrm{int}(X)$[45].
3. M is chaotic i.e. for its Lyapunov Exponent $\lambda = \ln(s)$ holds $\lambda > 0$.

In the context of modelling of economic price time series it is mandatory to be restricted only to diffusion processes with a non negative outcome. Therefore we shall specify another class $\Im_{>0}$ with the same properties like maps of $\Im$ with only positive values permitted.

Call $\Im_{>0}$ the family of piece wise linear maps $M_{>0}$: $\mathbb{R}^+ \rightarrow \mathbb{R}^+$ with uniform slope s having the properties:

1. There exists a map $M \in \Im$ with $M_{>0}=M$ if $\mu_{M,n} > 0$ with $\mu_{M,n} = \min(M([n-1,n]))$, $n \geq 1$.
2. If $\mu_{M,n} < 0$ $M_{>0}(X)= M-\mu_{M,n}$.
3.

The example maps $S \subset \Im$ and $S_{>0} \subset \Im_{>0}$ to be considered in the following are the saw tooth map S(X) and the on $\mathbb{R}^-$ restricted saw tooth map $S_{>0}(X)$.

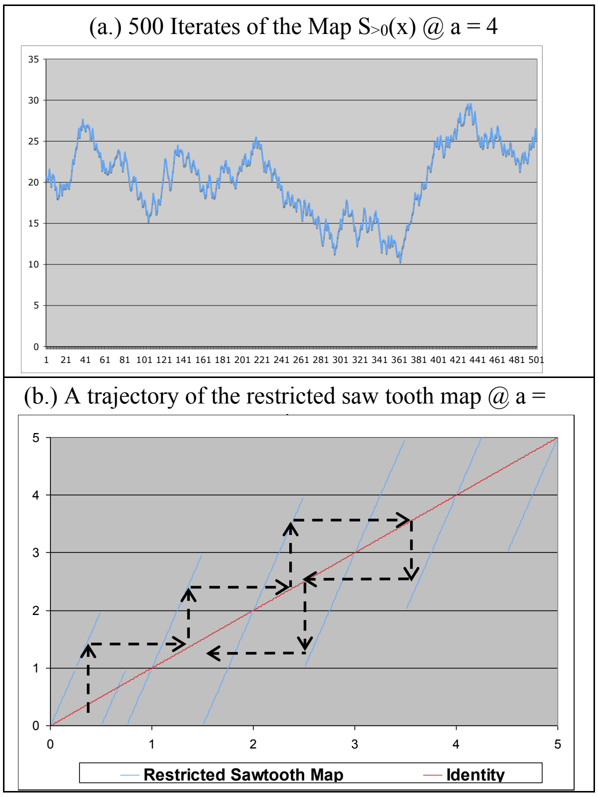


Figure 1. (a.) Iterates of the restricted Saw Tooth Map and (b.) a trajectory shown schematically.

S(X) is defined by:

$$S(X) = \begin{cases} s\left[x - \mathrm{int}(x)\right] + \mathrm{int}(x) \\ if \left\{ 0 < x - \mathrm{int}(x) < \frac{1}{2} \right\} \\ s\left[x - \mathrm{int}(x)\right] - s + 1 + \mathrm{int}(x) \\ if \left\{ \frac{1}{2} < x - \mathrm{int}(x) < 1 \right\} \end{cases} \quad (2)$$

[45] n is the smallest integer smaller than X

with s > 2. Note that for $s \in [1,2]$ S is chaotic but its iterates do not leave the unit interval. To yield its analogon $S_{>0}(X)$ in $\Im_{>0}$, S(X) needs to be modified to require S(X) >> 0.

$$S_{>0}(X) = \begin{cases} s\left[x - \mathrm{int}(x)\right] + \mathrm{int}(x) \\ if \ \left\{0 < x - \mathrm{int}(x) < \frac{1}{2}\right\} \\ s\left[x - \mathrm{int}(x)\right] - s + 1 + \mathrm{int}(x) \\ if \left\{\left(\frac{1}{2} < x - \mathrm{int}(x) < 1\right) \cup \left(s \le 2 + 2\,\mathrm{int}(x)\right)\right\} \\ s\left[x - \mathrm{int}(x)\right] - \frac{1}{2}s \\ if \left\{\left(\frac{1}{2} < x - \mathrm{int}(x) < 1\right) \cup \left(s > 2 + 2\,\mathrm{int}(x)\right)\right\} \end{cases} \tag{3}$$

To illustrate how the dynamics of S and $S_{>0}$ work in Figure 1(a.) a trajectory of the map $S_{>0}$ is shown schematically and as time series. As one can see, the behaviour of the map already resembles the behaviour of stock market prices. There are crashes and booms and periods of only slight movements. A more detailed introduction to Deterministic Diffusion can be found in [3] and [4].

3. Chaotic Stock Pricing

One motivation for the choice of Deterministic Diffusion as model for stock price movements is that it can be reproduced in an extended model framework of Day and Huang [5] presented in the following. The original model did not allow the stock price to diffuse. Therefore it will be enriched to permit the prices to show diffusion. Consider three phenotypes of investors:

1.) α-investors

The so called sophisticated investors have the following demand function:

$$D_\alpha\left(p, p_F^\alpha, d, u\right) = \left(a\left[p_F^\alpha - p\right]\right)\Theta\left(p, d, u\right), a > 0$$
$$\Theta\left(p, d, u\right) = \left(p - d + 0.01\right)^{-\delta_1}\left(u - p + 0.01\right)^{\delta_2} \tag{4}$$

Once the stock price is above the level they assume to be the fundamental price p_F^α according to some public information they want to sell the stock because they expect it to decline towards the fundamental value. On the other hand they buy the stock when it is cheaper than the fundamental price they assume. The strength of their reaction is determined by the parameter a > 0 and the chance that the stock price will fall or rise respectively if it is above or beneath p_F^α is expressed by the chance function Θ(p,d,u).

The chance of a price fall or rise will be judged by α-investors more likely the more the price is distinct from p_F^α. The demand of an investor is bounded by the levels u and d which represent the highest price he or she would sell and the lowest price he or she would buy. When prices are above u the chance of loosing money on a crash is perceived as too high therefore the asset is not bought. If the price is lower than d then the chance that it will ever rise again will be perceived as too little due to the fact that other investors in the market do not seem to be rational enough to allow for a reasonable stock pricing. The parameters δ_1 and δ_2 represent the relative strength of the bottoming or topping price d and u respectively in the chance function.

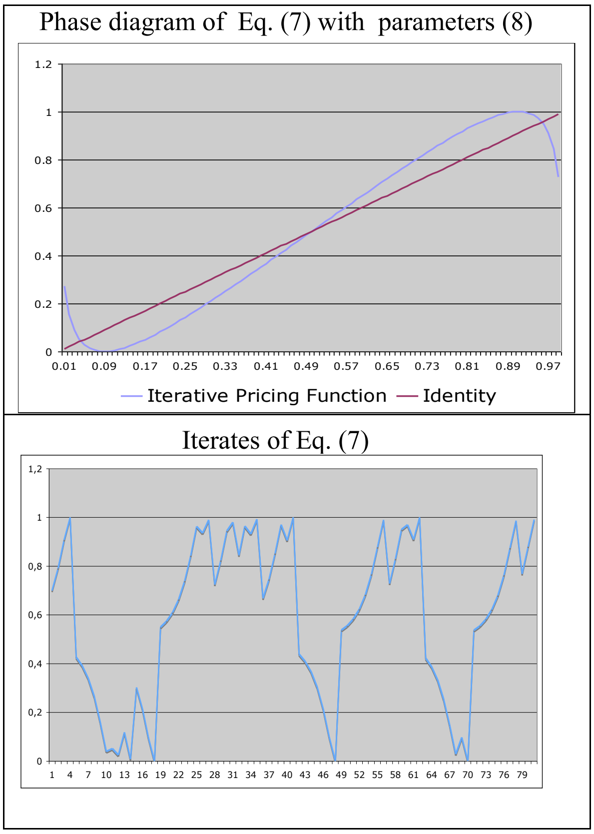


Figure 2. 1 Phase diagram and iterates of the iterative pricing equation (7) with parameters (8).

2.) β-investors

Are the less sophisticated investors. They expect the prices to rise when they are above the fundamental value and they estimate them to decline when they are beneath the fundamental value. One might call them trend-investors. Their demand function is represented as:

$$D_\beta(p, p_F^\beta) = b\left(p - p_F^\beta\right), b > 0 \qquad (5)$$

The parameter b > 0 represents the strength of their demand reacting to the difference between p and p_F^β.

3.) Investors

Simply make the market in the sense that they buy excess supply and sell from their own stock in case of excess demand E(p). In case of excess supply they lower the price and in case of excess demand they rise the price in order to sell or buy not too much from or to their inventory in line to keep their stock on a reasonable average over time. Their demand and price adjustment function θ is given as:

$$D_\gamma\left(p, E(p)\right) = -E(p)$$
$$\theta(p) = p + cE(p) \qquad (6)$$

Where c>0 is the adjustment parameter for the price.

In our simple dynamical model one yields the following iterative price formula Θ:

$$p_{t+1} = \theta(p_t) = p_t + c\begin{bmatrix} \left(a\left[p_F^{\alpha} - p_t\right]\right)\left(p_t - d + 0.01\right)^{-\delta_1} \\ \left(u - p_t + 0.01\right)^{-\delta_2} + b\left(p - p_F^{\beta}\right) \end{bmatrix} \quad (7)$$

In [5] the parameters for a numerical experiment were chosen to be:

$$d = 0, \delta_1 = \delta_2 = 0.5, u = 1, \\ a = 0.3, b = 0.88\, c = 2, \\ p_F^{\alpha} = p_F^{\beta} = p_F = 0.5 \quad (8)$$

Given this parameter setting the systems fixed points of every period (i.e. cycles and equilibrium prices) are unstable and dense in [d , u] thus deterministic chaotic motion is generated intrinsically by the model. See [5].

Figure 2 shows a trajectory and the phase diagram of the price adjustment equation Eq. (1). The trajectories do not look realistic and the price does not diffuse. To improve the model, some modifications will be considered.

Assume that not just three groups of investors are on the market but rather beside γ-Investors N different groups of α- and β-investors α_n, β_n: n=1,2,3,…..,N with the same parameters a, b ∀ n and with topping and bottoming prices u_n, d_n satisfying 1<n<N: $u_n=n*u_1$, $d_1=0$, $d_n=(n-1)*d_2$, and fundamental prices $p^{\alpha}_{F,n}=p^{\beta}_{F,n}=p_{F,n}$ with: $p_{F,n}=(n-1)+p_{F,1}$, $d_1<p_{F,1}<u_1$. Further claim that if $p\in[d_n,u_n]$ $\max(\Theta(p)) > u_n$ and $\min(\Theta(p)) < d_n$ whenever $\Theta(d_n) >> 0$ and $\Theta(u_n) << u_N$ and claim $0 \leq \Theta(p) \leq u_N$, $\Theta(p) \in \Im_{>0}$ if $N = \infty$, i.e. γ-Investors never make negative prices. Assume $|\Theta'(p)| > 1$ and that at price levels $p\in]u_{n-1},d_{n+1}[$ only the group of α_n, β_n, and γ investors commits trading. Finally let the parameters δ_1^n and δ_2^n of the chance functions of all the α-investors be zero.

On the basis of these assumptions one can construct a chain of chaotic maps that obey Deterministic Diffusion and the iterative pricing formula Θ(p) is a piecewise linear map in p (see Figure 3).

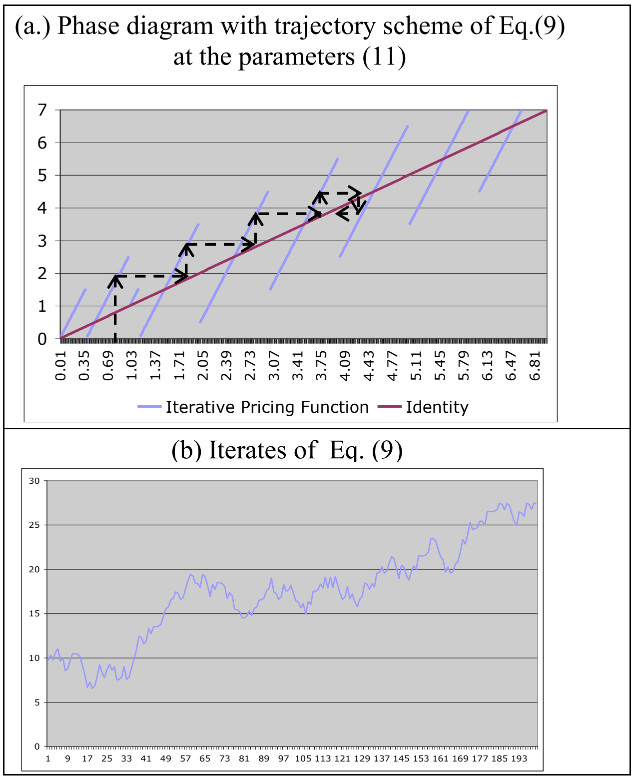


Figure 3. 1 Phase diagram with trajectory scheme and iterates of the iterative pricing equation (9) with parameters (11)

The model can now be expressed as:

$$p_{t+1} = \theta(p_t) = p_t + c\left[a\left(p_F^n - p_t\right) + b\left(p_t - p_F^n\right)\right]$$
$$if \begin{Bmatrix} \left(p_t + c\left[a\left(p_F^n - p_t\right) + b\left(p_t - p_F^n\right)\right] > 0\right) \cup \\ \left(p_t + c\left[a\left(p_F^n - p_t\right) + b\left(p_t - p_F^n\right)\right] < u_N\right) \end{Bmatrix} \quad (9.1)$$

$$p_{t+1} = \theta(p_t) = p_t + c\left[a\left(p_F^n - p_t\right) + b\left(p_t - p_F^n\right)\right]$$
$$-c\left[a\left(p_F^n - d_n\right) + b\left(d_n - p_F^n\right)\right] - d_n \quad (9.2)$$
$$if \left\{ p_t + c\left[a\left(p_F^n - p_t\right) + b\left(p_t - p_F^n\right)\right] < 0 \right\}$$

$$p_{t+1} = \theta(p_t) = p_t + c\left[a\left(p_F^n - p_t\right) + b\left(p_t - p_F^n\right)\right]$$
$$-c\left[a\left(p_F^n - u_N\right) + b\left(u_N - p_F^n\right)\right] - u_N \quad (9.3)$$
$$if \left\{ p_t + c\left[a\left(p_F^n - p_t\right) + b\left(p_t - p_F^n\right)\right] > u_N \right\}$$

With the requirement $int(p_t) = n-1$.

As can be seen, it is possible to derive an expectation driven asset pricing model with only a few assumptions regarding the pricing process and one can yield stochastic looking time series despite the fact are generated by a deterministic process. Note that the model permits rational and irrational behaviour as well as disagreement between investor groups regarding the fundamental value of an asset.

The time series of the model also exhibits patterns of real world stock price time series where sometimes a.) small changes are followed by small changes (no significant news/events) and b.) suddenly large changes are followed subsequently by large changes (stock market crash/boom). These patterns arise when: a.) α- and β-investors of one group n trade with γ-Investors and b.) suddenly $\Theta(p) > u_n$ or $\Theta(p) < d_n$ and the price gets adjusted so that the next group α- and β-investors involves in trading and this happens for a few subsequent groups in a row. Such behaviour is typical for stock market prices and hence an argument to use deterministic scattering maps to model stock market price movements.

In context of what has been presented we can conclude that expectations and behavioural patterns might drive the price in the context of Deterministic Diffusion and the behaviour of those artificial time series seems to mime the real world very well. The independence assumption of Gaussian white noise seems hence too restrictive and too naive. In general the stock market could presumably be better understood as a deterministic scattering mechanism where one event depends on the previous. Please note that the considered scattering maps can be understood as a simplified model of a poincare′ map of a deterministic system in one of its unstable directions. For details see [4].

4. Stylized facts of stock price time series

In this section the most popular empirically observed stylized facts of stock return distributions that are contradictory to the assumption of Brownian Motion are presented. Each fact gets exemplified with real world data of the German equity index DAX[TM] and the time series of the model from Section 3.

The stylized facts commonly observed on stock returns and their distributions are:

- **i.** Fat tails
- **ii.** Heteroscedasticity
- **iii.** Long range dependence

iv. Sensitivity to initial conditions

It will be shown that deterministic diffusive processes like the model of Section 3 have similar features and can therefore, in contrast to simple random walks, give better explanation to real world behaviour of stock prices. Subject to the forthcoming analysis in this section were 6800 model iterates and 6832 consecutive daily closing prices of the German equity index DAX^{TM} from 01.01.1976 to 22.02.2003. The parameter values of the model used for all following numerical investigations were:

$$d_1 = 0, u_1 = 1, a = 0.2 + 0.01 * \sqrt{2}, b = 1 + 0.01 * \sqrt{10}$$
$$c = 4,25 + 0.0001 * \sqrt{3}, p_F^{\alpha} = p_F^{\beta} = p_F = 0.5 \tag{11}$$

The choice of the irrational parameter settings Eq. (11) was motivated by the fact that most parameters in the real world should be irrational.

4.1 Fat tails

The tails of a probability distribution describe the probability of the occurrence of extreme events. They are called fat if this probability does not decay exponentially with the magnitude of the event. This feature has been observed frequently in stock market returns.

Definition 4.1.1 *(Fat tailed probability distribution)* A probability distribution P is called fat tailed if the probability of extreme events vanishes by a power law with an exponent α and the following scaling law holds:

$$P\left[|X| > x\right] \approx x^{-\alpha} \ \alpha > 0 \tag{10}$$

Note that if α < 2 the variance and all higher moments of the distribution do not exist. We will examine our model log returns by plotting log[P(|X|>x)] against –log(x). The slope of the regression curve is used as estimate for α. An extreme event was assumed to be at least two standard deviations away from the centre of the distribution. In Figure 4(a). the long term behaviour of the model at parameters Eq. (11) is shown for 2000 iterates. Additionally the volatility of 50 consecutive values is implemented in the same graph.

In Figure 4(b) the log return distribution of the model is compared to a standard normal distribution and finally Figure 4(c) shows log[P(|X|>x)] against –log(x) plot of the model. The slope estimated from this plotting was α ≈ 3.16. In figure 5(a)-(c) the same analysis with the DAX^{TM} time series is shown. The value of α for the DAX^{TM} time series was estimated to be 4.02. Both time series have fat tails and show strong heteroskedastic behaviour. From this section the reader should have got the first impression how fat tails and volatility clustering of stock returns could be related to a deterministic law generating them.

4.2 Heteroscedasticity

Heteroscedasticity is a common feature observed in stock market time series and also other economic time series. It happens to occur when a lot of large changes abruptly follow a series of moderate changes.

Definition 4.2.1*(Heteroscedasticity)*

Define the m-sample variance estimator at sample point k of a sample of N realizations of a variable $x_1, x_2, \ldots, x_N$ as:

$$\hat{\sigma}^2_{m,k} = \frac{1}{m-1} \sum_{i=k}^{m+k} \left(x_i - \hat{\mu}_{m,k} \right) \tag{12}$$

where $\hat{\mu}_{m,k} = \frac{1}{m} \sum_{i=k}^{m+k} x_i$ (13)

with 1<k<N and n+k<N ∀ k is defined as the m-sample mean estimator at sample point k. And define the sample variance by:

$$\hat{\sigma}_N^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \hat{\mu}_N\right) \qquad (14)$$

where $\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} x_i$ (15)

is defined as the sample mean estimator of the sample variance.

In this context heteroscedasticity would mean that for N samples there exists significantly many values k so that the m-sample estimators of those sample buckets differ significantly. From Figure 4(a.) and 5(a.) we can see clearly that the model generated price time series and DAX^{TM} should show heteroscedasticity. The word significantly can be given a meaning by a statistical hypothesis test as in [1] and [6].

4.3 Long Range Dependence

One striking feature of Brownian motion is that it has no memory. Thus all realizations are independent from one another in time. As will be shown shortly real world stock market returns and the model returns show exactly the opposite.

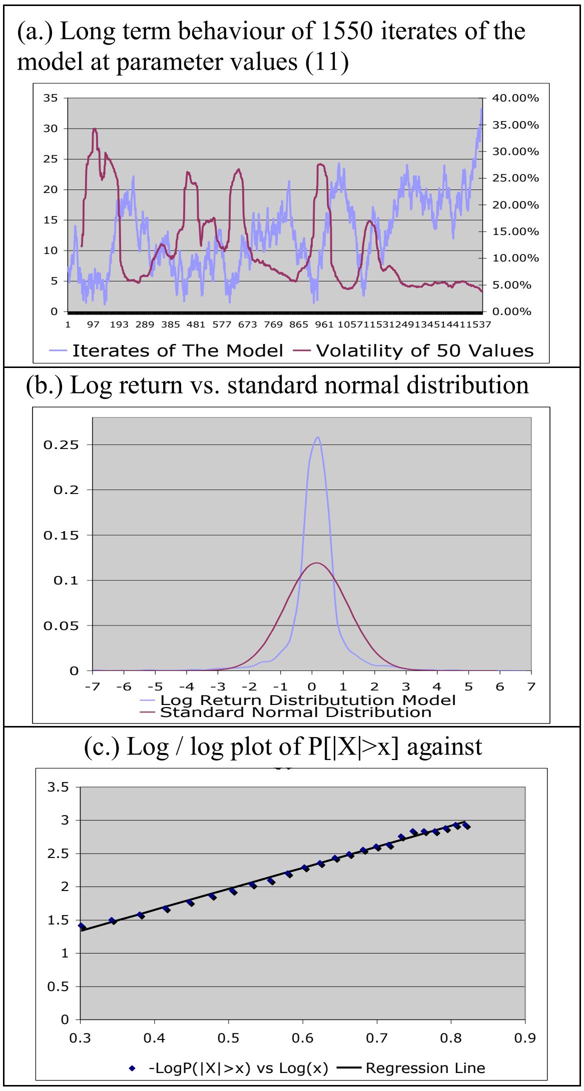


Figure 4: (a.) 1550 iterates of the model at parameter values Eq. (11) plus the volatility of 50 consecutive returns calculated on revolving time intervals. (b.) Comparison of log return distribution of the model against a standard normal distribution. (c.) Plot of log(P(|X|>x)) against log(X) for model returns.

Definition 4.3.1 *(Long range dependence)*

A time dependent process x(t) is said to be long range dependent, if the autocorrelation of its absolute time lagged values raised by any power k >=1 is greater than zero and decays in time by a power law with the rate δ^k.

$$\forall t > 0 s > 0, k >= 1:$$

$$\rho(s,k) = corr\left(|x(t)|^k, |x(t+s)|^k\right) > 0 \; and$$

$$\rho(s,k) \approx corr\left(|x(t)|^k, |x(t+1)|^k\right) s^{\delta_k} \qquad (16)$$

The lower the absolute value of δ^k the less the decay of dependence.

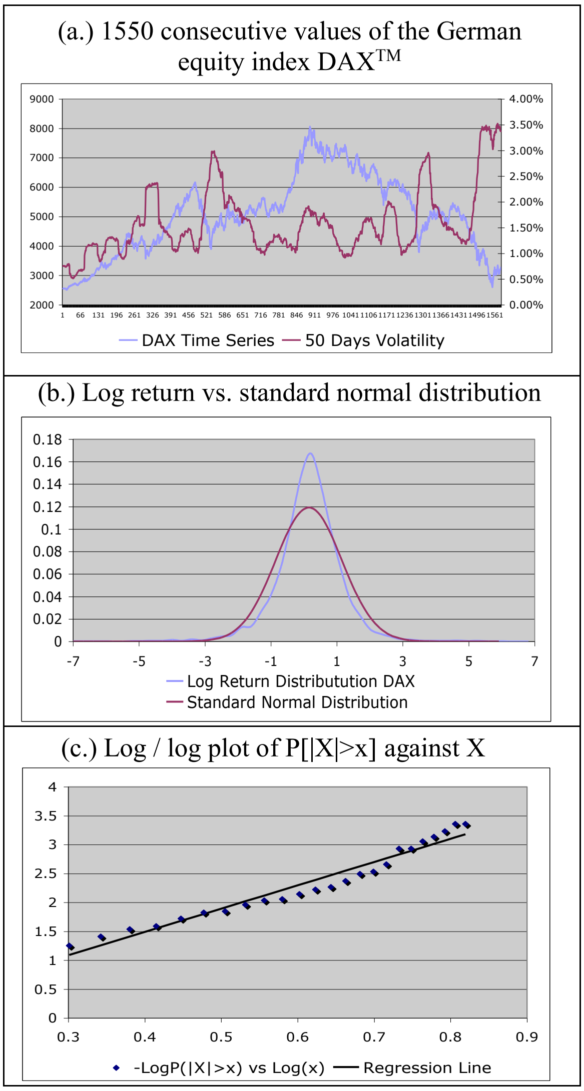


Figure 5. (a) 1550 iterates of the german equity index DAX plus the volatility of 50 consecutive returns calculated on revolving time intervals. (b) Comparison of log return distribution of the DAX™ time series against a standard normal distribution. (c) Plot log[P(|X|>x)] against log(X) for DAX™ returns.

To examine the DAX™ and the model time series of log returns ρ(n,2), $n \in \mathbb{R}^+$ was calculated as well as the ordinary autocorrelation function. The results are shown in Figure 6. Both time series show the same qualitative behaviour, but with different quantitative peculiarity.

The model autocorrelations of squared returns start at a very high level and decay very fast where as the real world time series correlations start at a lower level and decay more slowly. For the model series δ_2= -0.5 and for the DAX™ δ_2 = -0.3 got computed showing a faster decay and loss of dependence in the model time series than in the DAX™ returns.

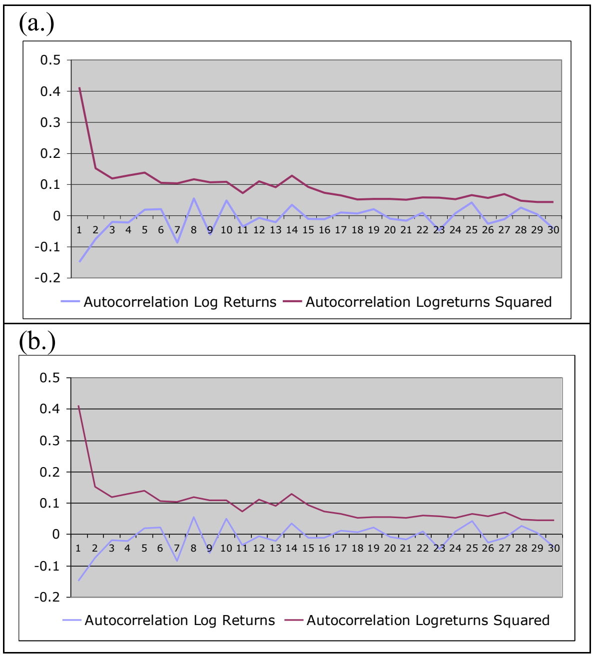


Figure 6. (a.) Autocorrelation Correlation Decay for normal and squared model time series of log returns with δ^2 =-0.5 (b.) Autocorrelation Correlation Decay for normal and squared DAX™ time series of log returns with δ^2 =-0.3.

Another measure of long range dependence or persistence is the Hurst Exponent, denoted in the following with H. It is named after its inventor, the hydrologist Harold Edwin Hurst. He invented it when analysing yearly water run offs of the Nile river. Consider n observations of a variable x: $x_1, x_2, x_3, \ldots\ldots, x_n$ and the cumulated values $X_k = x_1 + x_2 + x_3 + \ldots.\ x_k$. The value $X_k - (k/n)\ X_n$ measures the divergence of the cumulated value of a time series of length k from the rescaled cumulated value of the whole time series. Define the Range R_n as:

$$R_n = \max_{1\le k\le n}\left(X_k - \frac{k}{n}X_n\right) - \min_{1\le k\le n}\left(X_k - \frac{k}{n}X_n\right) \tag{17}$$

The empirical Standard deviation is given by:

$$S_n = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(x_k - \frac{X_n}{n}\right)^2} \tag{18}$$

Hurst found that for the rescaled range R/S:

$$\left(R/S\right)_n = \frac{R_n}{S_n} \approx cn^H \tag{19}$$

where the values X_k had approximately the same distribution like $n^H X_1$ with $H \approx 0.7$, indicating that the Nile River run-offs are not i.i.d. random events, but rather depend on one another persistently.

A process with such a scaling behaviour and distributional congru- ency is called statistically self similar.

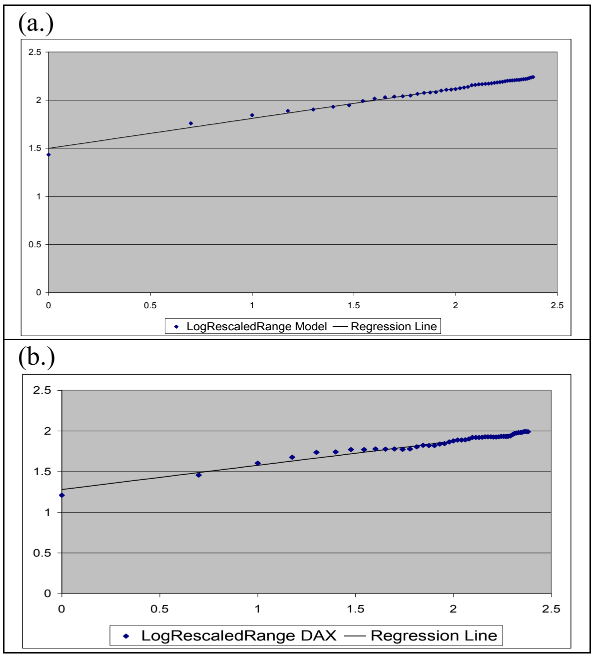


Figure 7. (a.) log/log plot for the R/S scaling of the model time series, slope estimated H= 0.31 (b.) log/log plot for the DAX^{TM} R/S scaling yielding H=0.29.

Definition 4.3.2 *(statistical self similarity)*
A statistical self similar process x is defined

by: $x(\alpha t) \cong x(t)\alpha^{2H}$ (20)

where x(t) is the value of the process after t time steps, $\alpha \in \mathbb{R}^{+}$, H∈[0,1] is the Hurst Exponent and the operator ≅ means congruency in distribution. If H > 0.5 a process is called persistent, If H < 0.5 a process is called anti-persistent for H=0.5 the process is called stable (since it's stable under addition).

The estimation of H from empirical data is straight forward. One plots log((R/S)n) against log(n). The slope of the regression line holds as estimate for H. Figure 7(a.) and (b.) show the resulting log/log plots for the model time series and the DAX^{TM} time series respectively.

The value H=0.31 was estimated for the model time series and H=0.29 for the DAX^{TM} time series. The DAX^{TM} and the model estimated Hurst Exponent values show clearly that both time series are anti-persistent and long range dependent. To get a better insight in this phenomenon, that should be called Hurst Effect in the following, Figure 8 shows the maximum and the minimum standardized log return of the DAX^{TM} and model time series relative to the time frame that was observed. The extreme events of both time series seem to decay by a power law in time. This is a commonly observed effect (called: mean reversion) in stock market time series. The loss or gain one experiences decreases with time, after periods of large losses new speculators enter the market and after periods of gains, positions get unwound in the course of realizing profits. Another way to look at this phenomenon is to observe the information entropy.

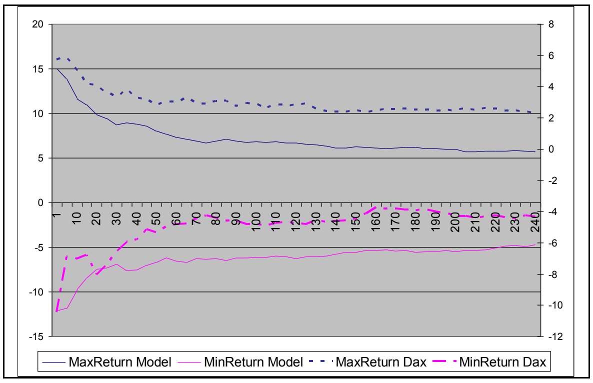


Figure 8: Max and Min standardized log returns of model (primary axis) and DAX™ time series (secondary axis).

Definition 4.3.2*(The Information Entropy)*

The Information Entropy for a random variable x with density h(x) according to Shannon is defined by:

$$\Pi = -\int h(x)\log_2\left(h(x)\right)dx \tag{21}$$

It gives the maximum Information in bits one learns from one outcome of a random variable. Thus the higher the information entropy, the more information produces an experiment. For example consider a coin toss. The information entropy of it is:

$$-2*(1/2)*\log_2(1/2) = -\log_2(1/2)=1 \tag{22}$$

Now consider a skew of the coin toss so that the probability of the one side of the coin turns ¼ and that of the other ¾. The information entropy then is:

$$-(1/4)*\log_2(1/4)- (3/4)*\log_2(3/4) = 0.81. \tag{23}$$

Thus the experiment needs to be repeated more often to get the same information than one outcome of the not skewed coin toss produces. Figure 9. shows the development of the information entropy over different time horizons for the model time series and the DAX™ time series computed for the normalized distributions of their log returns. From observing Figure 9 it is clear, that we cannot be dealing with a self similar stochastic process in both cases, because for such a process the information entropy would be constant in time. Figure 8 underlines this conclusion since constancy in time would also be expected to hold for the min and max log returns. It can also be seen from Figure 9 that the information entropy increases with time, which means, that the riskiness or uncertainty involved in the process decreases since one learns more about the world by one experiment on a longer time horizon.

From the previous observations naturally the question arises how the distribution of the processes may evolve over time. To investigate this issue the normalized distributions for the time buckets 1,20,240 were drawn into one graph shown in Figure 10. Again both time series obey the same peculiarities. On different time scales the distributions diverge from each other showing different shapes. The shape of the distribution of log returns depends on the time horizon, unlike the ones of a stable or self similar stochastic process.

As time progresses, autocorrelation of squares are strictly positive and decay slowly by a power law and risk declines measured in the form of the information entropy and min/max returns. Hurst analysis shows strong anti persistency for the DAX™ and the model. Contrary to a stable or self similar stochastic process the time dependent normalized frequency distributions of the DAX™ and model returns are not alike.

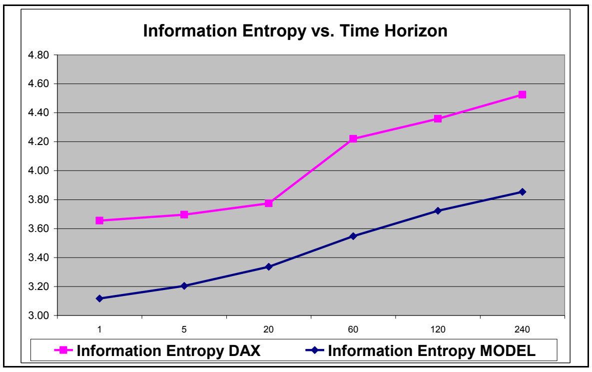


Figure 9. Development of the information entropy of model and DAX time series log returns in relation to time horizon in days/iterations.

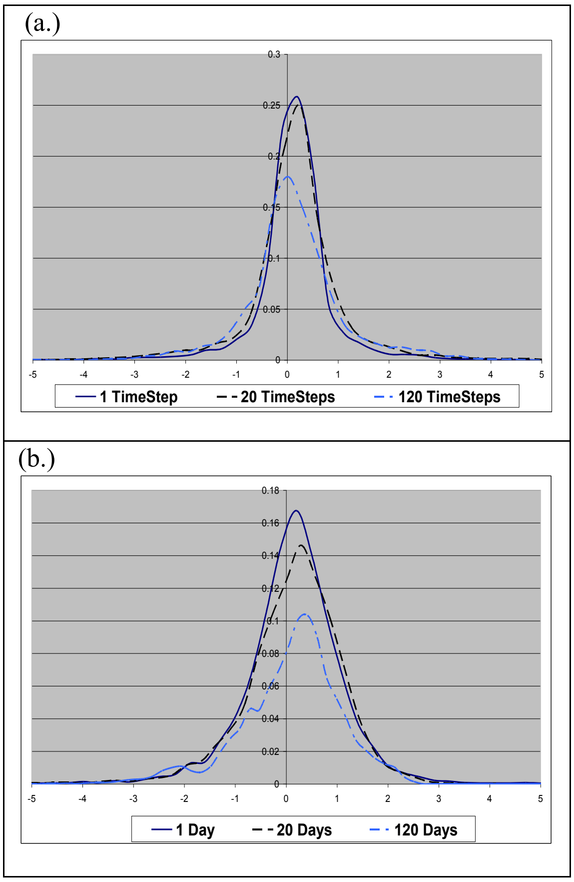


Figure 10. (a.) and (b.) the densities for the (a.) model and (b) DAX^{TM} log returns for different time horizons.

4.4 Sensitivity to initial conditions

Apart from the classical stylized facts the author would like to add this section to make the sections following thereafter more clear. Sensitivity to initial conditions is the property of a dynamic system that describes how it reacts on a small difference in the starting value in the long run. A popular measure of this kind of behaviour is the Lyapunov Exponent. It describes the average exponential expansion rate of a small error in the initial conditions.

Definition 4.1.1 *(Sensitivity to initial conditions)* A dynamical system is said to be sensitive to its initial conditions if a small error δx expands on the average exponential with rate $\lambda > 0$, called Lyapunov Exponent. The formal definition of λ is given by:

$$\lambda = \lim_{t \to \infty} \lim_{\delta x \to 0} \frac{1}{t} \sum_{n=1}^{t} \ln \left(\frac{|\delta x_n|}{|\delta x_0|} \right) \tag{24}$$

where δx_0 is the error in the initial condition of the iterates x(t) of the system and δx_t is the error after t time steps.

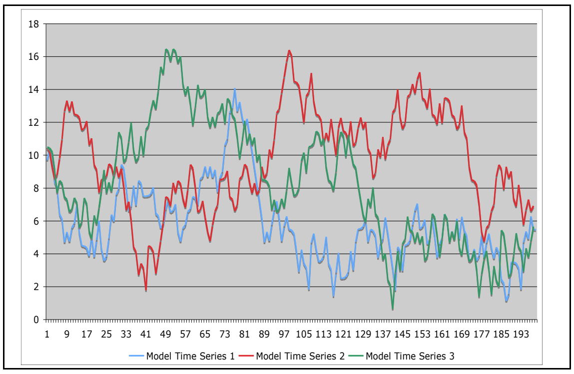


Figure 11 Model time series for initial conditions 9.7; 9.71; 9.72 respectively yielding extremely different trajectories.

Table 1 Results of the sensitivity analysis of the DAX™ and model time series

	Model	DAX
Lyapunov Exponent:	2.76	3.52
Lyapunov Time	3.08	3.46
initial Error	0.01	0.5
System Extend	50	10000

The meaning of this sensitivity is illustrated in Figure 11. Three trajectories of the model all only one hundredth i.e. 0.01 apart from each other in the starting value are shown, resulting in significantly different trajectories.

As measure of sensitivity on initial conditions, the Lyapunov Exponents for the model and the DAX™ have been computed. The following method was applied. Denote x_{NN} the nearest neighbour of the starting value x(0) of a time series in the sense of:

$$|x_{NN} - x(0)| = \inf |x(n) - x(0)| \; n = 1, 2, .., T \tag{25}$$

where T is the length of the time series. Then the following quantity holds as estimate for λ:

$$\tilde{\lambda} = \frac{1}{T - I_{NN}} \sum_{n=1}^{T - I_{NN}} \ln \left(\frac{|x(n) - x(I_{nn} + n)|}{|x_{NN} - x(0)|} \right) \tag{26}$$

where I_{NN} is the time index of the nearest neighbour value x_{NN}.

For the DAX™ time series $\lambda_{DAX} = 3.52$ was obtained and for the model $\lambda_{model} = 4.2$. Thus the model has less forecast ability than the DAX™. To illustrate this statement consider the Lyapunov time, defined as the maximal time a dynamical system can be forecasted:

$$T_\lambda = -\frac{1}{\lambda} \ln \left(\frac{\delta x_0}{\in} \right) \tag{27}$$

The variable $\in$ denotes the maximal extend of the system. The Lyapunov times of the DAX™ and the model respective were estimated assuming an initial error of 0.5 and 0.01 as well as an extend of 10000 and 50 respectively. The DAX™ time series had a Lyapunov time of $T_\lambda = 3.46$ days where as the model had $T_\lambda = 3.08$ iterates. The results are summarized in Table 1. It turns out, that if we knew the true model only a limited forecast of approx. 3 Days for the DAX[TM] and the model would be possible.

5. Implications on economic theory

In the following paragraphs the implications of the so far introduced model of Deterministic Diffusion will be discussed and reviewed against classical results of capital markets theory like the Market Efficiency Hypothesis, the CAPM and the Black Scholes option pricing formula.

5.1 Market Efficiency / CAPM

In the beginning the classical Market Efficiency Hypothesis will be recalled and afterwards the Deterministic Diffusion model will be related to it.

Market Efficiency Hypothesis (MEH)

The Markets are:

1. Semi efficient

If all information about price histories is contained in the prices.

2. Efficient

If they are semi efficient and all public information is contained in the prices.

3. Strong efficient

If they are efficient and all non public information is contained in the prices.

In strong efficient markets all prices are assumed to follow random walks of geometric Brownian motion in classical capital market theory, since only the occurrence of new information changes the price, and the price and its history do not contain any information about its future development. The consequence for capital asset pricing is that in equilibrium an asset i is priced accordingly that the expected excess return w.r.t. the risk free interest rate $E[r_i]- r_f$ can be expressed in terms of the expected excess return of a market portfolio M $E[r_m]- r_f$ times a beta factor $\beta_{i,M}$ of a stock. The asset's price therefore reflects the risk premium to be paid to an investor in equilibrium relative to the Market M:

$$E\left[r_i\right]-r_f=\beta_{iM}\left[E\left[r_M\right]-r_f\right],\ \beta_{iM}=\frac{Cov(r_i,r_M)}{\sigma_M^2} \tag{28}$$

The just stated equation forms the heart of the CAPM (Capital Asset Pricing Model).

There have been numerous articles and empirical investigations on whether the CAPM holds or not. The focus of the following will be rather a theoretical reasoning about the validity of the CAPM in the framework of Deterministic Diffusion.

In an environment of Deterministic Diffusion efficiency is not the general case but rather only one part of the story. Efficiency is present only for certain time frames, when crowd behaviour does not dominate price movements (Small changes are followed by small changes). In contrast if crowd behaviour dominates, as represented by β-Investors, (Large changes are followed by large changes), markets loose their efficiency, since trend movements set on and could be exploited by arbitrageurs. As can be seen from Figure 3(a) such a trend path triggered by succeeding entries or exits of several investor groups into the market is easy to construct in the model framework and should occur frequently. Still the question remains if such trends can be utilized systematically such that sustainable positive gains can be made. I.e. the information about the process presumably itself is hardly priced into a time series.

The MEH doesn't explain how the information that's priced into a stock is generated. If it is generated by a chaotic system and only the present state is responsible for the current price (short term pricing) then it would be no surprise that prices follow an unstable movement on a poincare′ map of this system.

Also the MEH doesn't allow for irrational and speculative investors like the β-Investors in our model. All investors should have the same opinion and information in the MEH and CAPM framework. This definitely does not hold in the real world, especially when it comes to the opinion about the future development of prices. It also should be obvious that limited rationality as well plays a significant role in the real world since the future is mostly unknown and impossible to be forecasted, especially on longer time horizons. Therefore no exact pricing of an asset may be possible and prices have to be revised daily.

To outline the different implications and assumptions in the preceding text the author would like to give a Chaotic Market Hypothesis CMH, that can be understood as modification or extension of the Fractal Market Hypothesis stated earlier by Peters in [8].

Chaotic Market Hypothesis (CMH)

1) Efficiency

Markets in general are not semi efficient in the sense of the MEH. There exist switching regimes between efficient and inefficient markets according to the strength of the action of β-Investors. Information gets incorporated into prices depending on the investor behaviour and sentiment. If it does, it reflects a long memory process of a large deterministic system. Every price is right as long as investors are willing to pay it. (The market is always right). The information about the price development process itself is hard to price. Arbitrage opportunities are rather of theoretical nature.

2) Investor Behaviour

Investors can act rational as well as irrational according to their personality or current sentiment. They have different investment horizons and different opinions about the future development of a price and have limited rationality i.e. limited knowledge about the future and the justified value of assets/shares.

3) Evolution of Prices

Prices diffuse according to deterministic laws, that can be to a certain extend interpreted as random. The diffusion is caused by a deterministic system driven by news and behavioural patterns of investors. The evolution law of prices shall be called "Deterministic Diffusion". It has infinite long memory and is not Markov.

When considering the CMH, the CAPM is only valid in a Deterministic Diffusion environment for short time horizons when markets can be interpreted as random walks, i.e. when markets are calm and efficient and the volatility is constant. There from we conclude that classical asset pricing models give a good understanding of how prices should be, but only capture certain aspects of real world stock market time series.

5.2 Option Pricing

At this point the author would like to sketch how an alternative option pricing model should look like.

In the classical investment theoretical framework the Black Scholes (BS) formula is well established. The major problem of this formula is the assumption of a homoscedastic log normal distribution for the price movements. The most crucial input parameter is the volatility since it is heteroscedastic in real economic time series. Any alternative option pricing model should therefore come up with a formula, that does not need the volatility as an input or can deal with heteroscedasticity. However there have been some alternative models brought up like the Option Pricing in case of fractional Brownian Motion, see [9] or Option Pricing for levy stable processes, see [10],[11]. In case of a fractional Brownian Motion a closed form solution exists.

A fractional Brownian Motion FBM stochastic process B^H is a Gaussian processes $B^H(t) \approx N(\sigma^H(t),\mu^H(t))$ with conditional second moment:

$$\sigma_H^2(t,T) = \left(T^{2H} - t^{2H}\right)\sigma^2 \; ; \mu_H(t) = B_H(t) \tag{29}$$

and its values are correlated by the covariance function:

$$E\left[B_H(t)B_H(s)\right] = \frac{1}{2}\left(|s|^{2H} + |t|^{2H} - |t-s|^{2H}\right)\sigma^2 \tag{30}$$

σ can be interpreted as the instantaneous volatility per H weighted unit of time. H is again the Hurst exponent.

The fractional Brownian Motion Black Scholes formula reads:

$$C(t,T,r,S(t),X) = S(t)N(d1) - Xe^{-r(T-t)}N(d2)$$
$$P(t,T,r,S(t),X) = Xe^{-r(T-t)}N(-d1) - S(t)N(-d2) \tag{31}$$

where:

$$d1 = \frac{\ln\left(\frac{S(t)}{K}\right) + r\left(T-t\right) + \frac{\sigma^2}{2}\left(T^{2H} - t^{2H}\right)}{\sigma\sqrt{T^{2H} - t^{2H}}} \tag{32}$$

$$d2 = \frac{\ln\left(\frac{S(t)}{K}\right) + r\left(T-t\right) - \frac{\sigma^2}{2}\left(T^{2H} - t^{2H}\right)}{\sigma\sqrt{T^{2H} - t^{2H}}} \tag{33}$$

C is the price of a European call option and P is the price of a European put option respectively. T > t is the date after the evaluation date t where the option matures, X is the strike price of the option, r is the risk free zero interest rate from the evaluation date until the date of maturity.

A random variable X is called α (or levy-) stable if and only if X ≅ γZ+δ, γ > 0 where Z has a characteristic function:

$$E\left[\exp(isZ)\right] = \left\{\begin{matrix} \exp\left(-|s|^{\alpha}\left[1 - i\beta\tan\frac{\pi\alpha}{2}\left(sign[s]\right)\right]\right) & \alpha \neq 1 \\ \exp\left(-|s|\left[1 + i\beta\frac{2}{\pi}\left(sign[s]\right)\log\left(|s|\right)\right]\right) & \alpha = 1 \end{matrix}\right\} \tag{34}$$

with α ∈ (0,2],β ∈ [-1,1]. Note that for α = 2 X is distributed with a Normal distribution that has volatility 2^0.5γ and expected value δ. Thus the Normal Distribution is a special case of Z with α = 2. The parameter δ is called the location parameter and is identical with the expected value of the distribution for 1 ≤ α ≤ 2. The parameter γ is called the scale parameter and is for α = 2 identical with half the variance of the distribution. The parameter β describes the skewness of the distribution. For β = -1 the distribution is skewed completely to the left and for β = 1 skewed completely to the right. The parameter α is called the characteristic exponent and describes the tail behaviour of the distribution

Beside those models from [9], [10] [11] define a progress in financial economics they still have certain drawbacks.

a. It has been shown in other papers that there exists arbitrage in FBM See [12].

b. The findings of this paper suggests, that equity price time series are neither a stable or self similar stochastic process like levy stable distributions or FBM are.

c. The assumptions made by Mc Cullogh in [10],[11] are too restrictive to apply.

d. The estimates of the tale parameter α in this paper are clearly above 2 while for levy stable distributions holds α < 2.

Definition 5.2.1 *(ergodic / invariant measure)*

A Map M: $\mathbb{R} \to \mathbb{R}$ is called ergodic if there exists a measure (density) 0 ≤ ρ(x) ≤ 1 such that:

$$\rho(x) = \lim_{n\to\infty}\frac{1}{n}\int_{\mathbb{R}}\sum_{0}^{n}\delta\left(M^{n}\left(x_0\right) - x\right)\rho\left(x_0\right)dx_0 \tag{35}$$

δ is the Dirac delta function, M^n denotes the n^{th} iterate of the map M, the integral is over all initial conditions x_0 in $\mathbb{R}$. The measure ρ(x) is called the invariant measure of M.

Note, since we are dealing with a deterministic process ρ is not a probability measure but rather measures the frequency of appearances of a value x in $\mathbb{R}$ on average. The kernel of (2.7) $\delta(M^n(x_0)-x)$ is called the Frobenius Perron Operator.

Definition 5.2.2 *(Conditional measure)*

The conditional measure $\rho_n(x \mid x_t;\varepsilon)$ given the state x_t of a Map M and measurement error ε of initial conditions is defined by:

$$\rho_n(x|x_t;\varepsilon) = \frac{1}{\rho\left(M^n(I)\right)}\int_I \delta\left(M^n(z)-x\right)\rho(z)dz \tag{36}$$

Where $I \in \mathbb{R}$ is an interval such that: $I = [x_t-\varepsilon/2;\ x_t+\varepsilon/2]$, ρ(x) is the invariant measure of the map and $\rho(M^n(I))$ is the "Mass" of the invariant measure on the n times iterated interval I by the Map M:

$$\rho\left(M^n(I)\right) = \int\limits_{M^n(I)} \delta(z)\rho(z)dz \tag{37}$$

Thus the conditional measure $\rho_n(x \mid x_t;\varepsilon)$ assigns a probability mass to every point $x \in \mathbb{R}$ that it can be reached by iterating the Map M, n steps forward in time given an initial state x_t and a measurement error ε.

Since ρ does not always exist as an applicable closed form formula, consider to partition I into k subintervals of length I/k. Furthermore assign the density 1/k to each point of I. Then a discrete approximation to (36) is:

$$\bar{\rho}_n(x|x_t;\varepsilon;k) = \frac{1}{k}\sum_{s=1}^{k}\chi\left(M^n\left(x_t+\frac{sI}{k}-\frac{\varepsilon}{2}\right)-x\right) \tag{38}$$

where χ is a characteristic function defined as: $\chi(0)=1 \ ; \ \chi(x<>0)=0$ (39)

Thus Eq. (36) can be computed at least by numerical simulations when the equations of driving the diffusive process are known. The derivation of a Statistic Dynamics (SD) option pricing formula is basically straight forward:

Theorem 5.2.1 *(Statistic Dynamics option pricing formula)* Given an ergodic Map M with invariant measure ρ(x), conditional measure $\rho_n(x \mid x_t;\varepsilon)$ given the state x_t of a Map M at time t and measurement error ε of initial conditions. An option with time to maturity T, and risk free interest rate r until maturity can be priced as:

$$Call\left(x_t,\varepsilon,T,r\right) = \int_0^\infty e^{-rT} Max\{S-X;0\}\rho_T\left(S|x_t;\varepsilon\right)dS \tag{40}$$

$$Put\left(x_t,\varepsilon,T,r\right) = \int_0^\infty e^{-rT} Max\{X-S;0\}\rho_T\left(S|x_t;\varepsilon\right)dS \tag{41}$$

Assuming risk neutral individuals, that can compute $\rho_n(x \mid x_t;\varepsilon)$ approximately.

To give the reader an idea on how Eq.(40) and (41) influence the option price, $\rho_n(x \mid x_t;\varepsilon)$ was computed by numerical simulations using the approximation Eq. (38). Subject of the simulation were 2000 iterates of the model. An approximation error of ε=0.01 was assumed, equivalent to that an accurate measurement only possible up to one hundredth. The interval I=1 was partitioned into one hundred equal long pieces of length 0.01 each being assigned a density of 0.01, the density was then evolved by numerical simulation until time to maturity and Theorem 5.2.1 was used to price the option.

Additional to the Statistic Dynamics option price the FBM price and the traditional Black Scholes price was computed. Two experiments were made. In the first experiment shown in Table 2 only the first 250 iterates were used in the estimation of the volatility. In the second experiment shown in Table 3 the volatility was estimated using all 2000 iterates. The Hurst Exponent was estimated always using all iterates (with value H=0.32) to give a better estimation quality. The two experiments were chosen to find out about the impact of the parameter estimation error that occurs when calibrating a homoscedastic model in a hetero- scedastic environment. In both numerical investigations the risk free interest rate r was assumed to be zero. The results of the experiments can be summarized as follows:

1) Heteroscedasticity Effect

Firstly the use of FBM and BS models for the pricing of options with the overall volatility of 9.8% is generally superior in comparison to use them with the volatility estimator of 26% of only the first 250 iterates. In particular out of the money options are extremely overpriced by the BS model with 26% volatility. The same holds for the FBM Model in a more moderate fashion. With 26% Volatility per time step a longer time to maturity has diminished influence on the BS option price. One can even see from table 2, that in the limit of infinite volatility and time to maturity the BS call price converges towards the spot and the BS put price converges towards the strike. The bias of the BS option price caused by heteroscedasticity is thus not negligible. The FMB option price has similar features, but with less strength since the presence of the Hurst Effect counter affects the heteroscedasticity bias.

2) Hurst Effect

Secondly the low Hurst Exponent of 0,32 indicates a strong anti-persistence effect that causes the log price not to diffuse unbounded, but to return to where it has come from (Recurrence). That's the reason why out of the money options are priced to high by the BS model in the Deterministic Diffusion environ- ment. The value of SD options decays rapidly to zero with decreasing moneyness. The FBM option prices also decay and match those of the SD option prices far better than the BS prices, since the FBM model captures the Hurst Effect of anti persistency.

3) Time to Maturity (Theta) Effect

Thirdly the larger the time to maturity (Theta), the larger the divergence to the pricing of the BS and FBM model of the SD pricing dependent on the heteroscedasticity bias. This goes well in line with the findings of section 4.3 that the process of Deterministic Diffusion does not posses a statistic self similarity and that the risk decreases on a long time horizon.

It can be concluded that the BS Model can lead to large price deviations from the SD price when the underlying price driving process is not the assumed process. The effect becomes more pronounced as the time evolves. The FBM Option pricing model gives a good approximation to the SD option prices, but needs a large sample of the time series to be estimated efficiently.

Table 2. Option Prices Experiment 2, Volatility = 26% of first 250 iterates was used, H=0.32 of total time series was used, Spot = 12.15

	Maturity = 30 time steps					
	Statistical Dynamics		Black Scholes		Fractional BM	
Strike	Call	Put	Call	Put	Call	Put
2	9,91	0,02	10,45	0,30	10,16	0,01
5	7,04	0,15	8,76	1,61	7,52	0,37
10	2,82	0,92	6,98	4,83	4,54	2,39
15	0,46	3,56	5,82	8,67	2,84	5,69
20	0,00	8,11	4,99	12,84	1,85	9,70
25	0,00	13,11	4,36	17,21	1,25	14,10
30	0,00	18,11	3,86	21,71	0,87	18,72
40	0,00	28,11	3,12	30,97	0,45	28,30
50	0,00	38,11	2,60	40,45	0,25	38,10
	Maturity = 100 time steps					
	Statistical Dynamics		Black Scholes		Fractional BM	
Strike	Call	PuDG	Call	Put	Call	Put
2	10,34	0,03	11,35	1,20	10,27	0,12
5	7,66	0,34	10,73	3,58	8,17	1,02
10	4,03	1,74	10,08	7,93	5,94	3,79
15	1,77	4,45	9,61	12,46	4,56	7,41
20	0,46	8,20	9,25	17,10	3,62	11,47
25	0,08	12,80	8,94	21,79	2,95	15,80
30	0,01	17,69	8,68	26,53	2,45	20,30
40	0,00	27,69	8,25	36,10	1,77	29,62
50	0,00	37,69	7,90	45,75	1,33	39,18

	Maturity = 250 time steps					
	Statistical Dynamics		Black Scholes		Fractional BM	
Strike	Call	PuDG	Call	Put	Call	Put
2	12,16	0,11	11,98	1,83	10,52	0,37
5	9,66	0,62	11,86	4,71	8,95	1,80
10	6,28	2,29	11,74	9,59	7,30	5,15
15	3,81	4,85	11,64	14,49	6,22	9,07
20	2,18	8,18	11,57	19,42	5,43	13,28
25	1,18	12,12	11,50	24,35	4,82	17,67
30	0,67	16,60	11,45	29,30	4,33	22,18
40	0,15	26,13	11,35	39,20	3,60	31,45
50	0,03	36,02	11,26	49,11	3,07	40,92

Of course in practise the approach presented here is not easily applicable because one normally does not know the true process driving the dynamics or the invariant measure of the stock price. But there should be possible methods yielding approximately comparable results. This should be part of future research.

6. Summary and conclusions

In the foregoing paper a new model named "Deterministic Diffusion" was introduced to model stock price processes. The model can be motivated by simple behavioural models of the stock market and does not need too many restrictive assumptions to be reasonable. Furthermore it helps understanding on how randomness comes about and how typical stylized facts like i.) heteroscedasticity ii.) long range dependency iii.) fat tailed frequency distributions in real world stock market data can be explained.

Table 3. Option Prices Experiment 2, Volatility = 9.8% of total 2000 iterates was used, H=0.32 of total time series was used Spot, = 11.26

	Maturity = 30 time steps					
	Statistical Dynamics		Black Scholes		Fractional BM	
Strike	Call	Put	Call	Put	Call	Put
2	9,91	0,02	10,15	0,00	10,15	0,00
5	7,04	0,15	7,24	0,09	7,15	0,00
10	2,82	0,92	3,58	1,43	2,64	0,49
15	0,46	3,56	1,68	4,53	0,54	3,39
20	0,00	8,11	0,79	8,64	0,08	7,93
25	0,00	13,11	0,39	13,24	0,01	12,86
30	0,00	18,11	0,20	18,05	0,00	17,85
40	0,00	28,11	0,06	27,91	0,00	27,85
50	0,00	38,11	0,02	37,87	0,00	37,85
	Maturity = 100 time steps					
	Statistical Dynamics		Black Scholes		Fractional BM	
Strike	Call	Put	Call	Put	Call	Put
2	10,34	0,03	10,21	0,06	10,15	0,00
5	7,66	0,34	7,87	0,72	7,17	0,02
10	4,03	1,74	5,34	3,19	3,15	1,00
15	1,77	4,45	3,81	6,66	1,15	4,00
20	0,46	8,20	2,83	10,68	0,40	8,25
25	0,08	12,80	2,17	15,02	0,14	12,99
30	0,01	17,69	1,70	19,55	0,05	17,90
40	0,00	27,69	1,10	28,95	0,01	27,86
50	0,00	37,69	0,76	38,61	0,00	37,85

	Maturity = 250 time steps					
	Statistical Dynamics		Black Scholes		Fractional BM	
Strike	Call	Put	Call	Put	Call	Put
2	12,16	0,11	10,53	0,38	10,15	0,00
5	9,66	0,62	8,99	1,84	7,27	0,12
10	6,28	2,29	7,36	5,21	3,73	1,58
15	3,81	4,85	6,29	9,14	1,85	4,70
20	2,18	8,18	5,51	13,36	0,94	8,79
25	1,18	12,12	4,91	17,76	0,49	13,34
30	0,67	16,60	4,43	22,28	0,27	18,12
40	0,15	26,13	3,70	31,55	0,09	27,94
50	0,03	36,02	3,17	41,02	0,03	37,88

Comparisons throughout the paper to real world DAX^{TM} time series show obvious parallel features that are not neglectable and give evidence for the appropriateness of the approach. Both time series have fat tailed frequency distributions of their log returns, slowly decaying autocorrelations of their squared returns and show a large degree of heteroscedasticity.

In a Hurst analysis for the Model time series and the DAX^{TM} time series showed strong anti-persistency with a value of $H \approx 0{,}3$. Log returns revert back to their mean and do not grow linearly in time since new speculators may enter the market any time or firstly invested speculators may unwind their positions from time to time to realize profits. By looking at frequency distributions on different time scales and the development of the information entropy in time both processes seem not to follow a simple self similar or stable stochastic law. An empirical estimation of Lyapunov Exponents results in clearly positive values for the DAX^{TM} and the model time series giving strong indication for the presence of deterministic chaos. Markets do have inefficient phases where behavioural patterns of investors dominate the price evolution.

Future research should be concerned with a.) A more precise description of the Deterministic Diffusion process in terms of variables or scattering maps with empirical fitting methods to existing time series, b.) Empirical detection methods of Deterministic Diffusion and c.) Easily applicable option pricing formulas and / or methods.

Furthermore other economic time series like exchange rates and interest rates could be analyzed to see if the Deterministic Diffusion model would be reasonable for them as well. The implications for Risk Management surely should also not be out of the scope of further investigations.

7. References

[1] Andrew W. Lo, and MacKinlay, A. Craig, (1987), *Stock Market Prices do not follow Random Walks: evidence from a simple specification Test*, Department of Finance, Wharton School, University of Pennsylvania, Philadelphia, PA 19104 USA

[2] Hans Malmsten, Timo Teräsvirta, (2004), *Stylized Facts of Financial Time Series and Three popular Models of Volatility*, Department of Economic Statistics Stokholm School of Economics Sweden.

[3] Heinz Georg Schuster, (1995), *Deterministic Chaos. An Introduction*, Wiley-VCH.

[4] *Chaos Book* various authors see http://www.chaosbook.org

[5] Richard H. Day, Huang Weihong, (1995), *Bulls, Bears and Market Sheep*, in: *Journal of Economic Behaviour and Organization*, Vol 12, pp 299 – 329.

[6] Guido Venier, (2008), *A simple Hypothesis Test for Heteroscedasticity*, (Draft).

[7] Nualart, D., (2006), *Stochastic calculus with respect to fractional Brownian motion*, in: *Annales de la faculté des sciences de Toulouse,* Sér. 6, 15 no. 1, p. 63 – 78.

[8] Edgar E. Peters, (1994), *Fractal MarketAnalysis*, (John Wiley and Sons, New York, NY.

[9] Necula, C., (2002), *Option Pricing in a Fractional Brownian Motion Enviroment*, Working paper of the Adademy of Economic Studies, Bucharest (Romania) 27, 8079{8089.

[10] McCulloch, J. Huston, (2003), *The Risk-Neutral Measure and Option Pricing under Log-Stable Uncertainty*, June 27.

[11] McCulloch, J. Huston, (1996a), *Financial Applications of Stable Distributions*, in: *Statistical Methods in Finance*, ed. by G.S. Maddala and C.R. Rao, Vol. 14 in: *Handbook of Statistics*. Amsterdam: Elsevier, pp. 393 – 425.

[12] Bender Ch., Sottinen T., and Valkeila E., (2006), *Theory of Stochastic Processes*, Vol.12 (28), no.3-4, 2006, pp.*-*, *Arbitrage with fractional Brownian motion?*.