Cracow University of Economics

ARGUMENTA OECONOMICA CRACOVIENSIA

No 15

Cracow 2016

Cracow University of Economics

ARGUMENTA OECONOMICA CRACOVIENSIA

No 15

Cracow 2016

ISSN 1642-168X

ISSN 1642-168X

Original version: printed publication
All the articles published in Argumenta Oeconomica Cracoviensia are available on the journal's website (www.aoc.uek.krakow.pl) and in the CEEOL database (http://www.ceeol.com)

Cracow University of Economics Press
Rakowicka 27, 31-510 Kraków, Poland
www.uek.krakow.pl

CONTENTS

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 15 • 2016
ISSN 1642-168X
AOC, 2016; 15: 5–7

Dear Readers,

The latest edition of *Argumenta Oeconomica Cracoviensia* confirms the broad range of subjects tackled in the journal, in which we publish articles from the fields of economics, management science, finance, and other sub-disciplines of economics. This broad profile creates publication opportunities for many authors working in these areas, which is reflected in the range of articles sent in to the editorial board.

Our aim is to publish both theoretical texts as well as articles that present the results of empirical research on macroeconomic and microeconomic phenomena at the national and international levels.

The issue of deepening income inequalities has been the subject of lively debate among economists, sociologists, and politicians in recent years. One of the views they have expressed is that this trend – if it continues – could present a serious risk to global economic growth. While there are a number of factors influencing the rise in income and wealth inequalities, there is general agreement that globalisation and the liberalisation of economic activity are having a powerful impact on income distribution within societies. These trends find particular expression in the location of production in more weakly developed economies, where labour costs are considerably lower. This causes revenues and profits to flow between countries, thus aggravating inequality at the international level. With the support of empirical research, James K. Galbraith addresses this intriguing issue in the opening paper, "Globalization and Inequality Revisited".

The workings of the Eurozone and its future prospects are a central concern for economists and politicians. This is understandable given that creation of the Eurozone may still be regarded as a huge economic and social experiment. The fate not only of the further monetary, but also social and political integration of the European Union may depend on the success or failure of the euro area. It follows that the results of research on the operation of the Eurozone – in the context of the established paradigms of contemporary banking and with reference to the national banking sector – also merit attention. Dependencies between deposits and lending in the banking sector emerge within the framework of these paradigms. In her article, "The Nature and Determinants of High-value Household Deposits in the Euro Area", Katarzyna Kochaniak undertakes a thought-provoking exposition of the role of deposits in the operation of the banking sector in the euro area.

"The Logic of Imitative Processes: Imitation as Secondary Innovation – An Axiomatic Schumpeterian Analysis", by Beata Ciałowicz and Andrzej Malawski, affirms the journal's declared policy of making room for papers that tackle theoretical and abstract issues. From the perspective of the search for innovative solutions, this article successfully presents a formalised conceptualisation of the hypothetical, creative behaviours of people in the operation of an economic entity. The active and creative attitude of the people involved is understood, within the terms of a Schumpeterian analysis, as constituting an opportunity for the organisation to develop.

The criteria for selecting features and the method used to analyse them are of vital importance in empirical investigations involving the assessment of observed entities (subjects). This issue is addressed by Sabina Denkowska in "Assessing the Robustness to Unobserved Confounders of the Average Treatment Effect on Treated Estimated by Propensity Score Matching". The value of the article lies in its verification of the Propensity Score Matching method, which will be familiar from the literature. In essence, this involves taking account of the observable and unobservable features of the subjects (entities) under analysis. The author's findings may help evaluate the usefulness of this method, which is important to the extent that it is recommended by the European Union as part of the procedure for accepting programmes for financing and implementation.

The search for the best (most efficient) method of building a portfolio of financial assets has been the subject of numerous attempts to formulate and verify theoretical constructs. It will therefore come as no surprise that the question has also been taken up in the pages of this journal: in this case by Przemysław Jaśko in his paper "Statistical Arbitrage: A Critical View". The theory of arbitrage occupies a prominent position in processes of portfolio building. For this reason, it may be assumed that a further attempt to critically assess cointegration tests used in the search for the log--price processes of related instruments when building a statistical arbitrage portfolio will prove of interest to readers.

The demographic processes taking place across the world – especially those in wealthy countries where societies are ageing due to civilisational progress and advancements in medical science – have been attracting increased interest not only from demographers but also from economists and politicians. It is for these reasons that readers may be interested in Grażyna Trzpiot and Justyna Majewska's article "Modelling Longevity Risk in the Context of Central Statistical Office Population Projections for Poland to 2050". With reference to demographic processes in a global perspective, the authors focus their research efforts on Poland and set out the risks associated with demographic change. The paper therefore takes its place alongside other Polish studies (Central Statistical Office, Social Insurance Agency) that have attempted to understand the threats to the country in the period up to 2050.

The present issue concludes with "R&D: Italy and Poland Compared" by Valerio Fino and Janusz Rosiek, in which the authors critically assess the workings of the R&D systems of these two countries. Their assessment, which is carried out by means of a SWOT analysis, may serve as a platform for possible changes in this field. This is especially true of Poland, where the system for implementing scientific achievements into industrial practice is ineffective and is evaluated as such by the authors. Their conclusion is that countries undergoing economic transformation are still looking for the best ways to organise, manage, and finance research and development. The authors suggest that certain Italian solutions could be adapted for Poland and for other countries with economies in transition.

Whilst commending the present issue to our readers, we also invite contributions in the form of original texts, information about important academic events, and reviews of outstanding books.

Prof. Stanisław Owsiak
Editor-in-chief

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 15 • 2016
ISSN 1642-168X
AOC, 2016; 15: 9–19
DOI: 10.15678/AOC.2016.1501

James K. Galbraith

GLOBALIZATION AND INEQUALITY REVISITED*

Abstract

This article summarizes the literature on inequality and economic development from Kuznets (1955) through the neo-classical labor-market-adjustment models and technology/education dialectics of the 1990s, and onward to the modern use of comprehensive, consistent data sets with global coverage to assess the interdependent and divergent experience of advanced, developing and transition economies in the age of globalization. Work based on the data sets developed by the University of Texas Inequality Project broadly validates Kuznets' original view of the importance of inter-sectoral transitions, but with many distinct and new insights, as it becomes possible to track regional and transnational patterns, common global macroeconomic forces and, most recently, the critical role played by exchange rates in the evolution of inequality in open economies.

Keywords: globalization, inequality, wages, labor markets, productivity.
JEL Classification: D63, E23, F60, J31.

In 1955, Simon Kuznets offered an intuitive account of the evolution of economic inequalities in the process of industrialization. At first, urban centers and factories would attract labor from the poor-but-egalitarian countryside, and the differential necessary to achieve this would be, for a time, the single most important source of inequality in the system. As cities grew, so too would inequality, until such time as the countryside mechanized and emptied out, and the now-industrial nation became predominantly citified. Then, Kuznets reasoned, inequalities would begin to fall – a process

James K. Galbraith, The University of Texas at Austin, Austin, Texas 78713-8925, SRH 2.237, USA, e-mail: galbraith@mail.utexas.edu

* This paper is lightly updated from a contribution entitled "Globalization and Inequality", first made in 2012 to the *Oxford Companion to International Relations*.

reinforced by the development of trade unions, democratic governance, and a redistributive welfare state.

Kuznets had very little evidence to work with, apart from the apparent experience of the United States, Great Britain, and a few nations of continental Europe through the middle of the twentieth century. In much of the world, moreover, the process of industrial development had barely begun, and in important parts of it, the Kuznets process had been short--circuited by communist revolution. Moreover, Kuznets had to be careful to restrict himself to inequalities of pay, excluding land rent from the picture; otherwise the starting point of (say) the antebellum South in the United States would have to be accounted very differently. Still, his was a powerful common-sense argument, and it dovetailed nicely with the optimistic liberalism of American development theorists at the time. So the Kuznets Curve – an inverted-U relationship between national income and economic inequality – became the archetype of hypothetical global-inequality relationships.

Over time, there was even a tendency for the original logic of Kuznets' argument to recede in memory, and for analysts to focus on finding an empirical inverted-U in comparative and historical data. As sources of data multiplied, this became increasingly difficult, for the apparent signal – if it existed at all – could not be reliably found in the noisy survey records on which researchers were prone to rely. For much of the post-war period, researchers using survey records could not even establish trend changes in inequality for most countries, including the United States. With no apparent trends, and the rising fashion of general-equilibrium models, the topic of distribution fell into obscurity. As a result, while Kuznets' curve remained an archetype, it ceased to command wide acceptance.

Within the United States, researchers began to notice that inequality was beginning to rise in the mid- to late 1980s; credit for calling attention to the phenomenon belongs in part to Bluestone and Harrison (1990), who offered an explicitly political and institutional argument, relating the rise of inequality to the consequences of the economic policies adopted under President Ronald Reagan, especially that administration's attack on trade unions. The mainstream of the economics profession took a different view, however, with two competing market-based arguments. One of these emphasized the role of technology, the other the role of trade.

Bound and Johnson (1992) proposed that rising wage inequalities were the result of increasing relative demand for workers with skills suitable to the changing requirements of employers; thus, "skill-biased technological change" entered the lexicon. Underlying this concept was the implicit view that changing inequality was driven by changing wage rates, reflecting changing marginal productivity, and thus that wage rates were set in an efficient and competitive labor market. The skill-bias hypothesis thus reinforced the view that the US labor market owed its favorable record on job creation and unemployment (compared to Europe) to its "flexibility" – to the latitude enjoyed by employers to match wage rates to the requirements of technology and the distribution of skills.

The concept of inequality driven by skill-biased technological change focused attention on the relative demand for skilled labor. An early alternative proposed that the true cause was instead a large increase in the effective supply of unskilled labor, due to the globalization of manufacturing and to immigration from developing countries. Thus the hypothesis of a rightward shift in relative demand for skills was countered by the hypothesis of a leftward shift in relative supply. Either hypothesis would produce a rising gap between skilled and unskilled pay rates; the distinction between them would turn on the effects on rates of employment among the skilled and the unskilled. The trade hypothesis also raised welfare and policy questions – whether the gains from more efficient world productive capacity justified the losses imposed on unskilled workers in the developed countries, and whether those losses deserved compensation. Still, like the skill-bias argument, the trade argument was built on the neoclassical foundation of efficient labor markets and marginal productivity wage setting, with an admixture of the Stolper-Samuelson relative-wage-equalization theorem.

Both variations on the labor-market-adjustment narrative generated empirical assertions that would prove problematic. The skill-bias hypothesis notably raised the question: what technical change? The obvious candidate was computerization, but the timing of the diffusion of personal computers came too late to account for rising inequalities (Galbraith 1998), and case studies of the effects of computerization on labor market outcomes failed to support the hypothesis (e.g., DiNardo & Pischke 1996). Meanwhile, the Stolper-Samuelson theorem appeared to predict that wage inequalities would decline in industrializing countries (Wood 1994) as they rose in developed countries, but concrete evidence for this effect proved hard to find. Moreover, the scale of increased trade and outsourcing could only with difficulty be stretched to account for the observed increases in inequality, and for their appearance in the non-traded-goods sectors.

As this debate developed, a third perspective (Baker et al. 2005) continued to make the institutional and political argument that the rise of inequality in the US was because of the decline of trade unions, the retreat of progressive political forces in general, and the declining real value of the minimum wage. While there was clear historical evidence for this view, it comported poorly with the labor-market models in fashion among economists, since it implied that wage rates had been political and administrative all along. If this were true, the role of economic theory in wage setting is easily dispensed with.

A difficulty with all three lines of argument was that all of them depended on changes in relative wage rates, measured on an hourly basis. This is what corresponds to the theoretical construct of marginal productivity, and what should vary in an "efficient labor market". Yet for practical purposes, no such measures exist. The actual information sets generally measured weekly earnings, with average hourly earnings derived by dividing hours into earnings. This procedure does not provide a reliable measure of hourly wage rates, however, since (thanks to overtime and other factors) earnings per hour vary with hours worked. Further, changing job classifications and job structures made it very difficult to assess whether rising inequality was due mainly to erosion or improvement of relative wage rates, as against changing structures of employment within or among firms. Indeed, the more one tried to isolate the effect of changes in relative hourly wage rates on overall wage inequality, the less important this factor appeared to be, as compared with changing patterns of employment, a changing composition of the workforce, and (for household earnings) a changing pattern of household composition and family life. Of all these forces, the hourly wage rate associated with any particular line of work appears to be one of the most stable.

Ferguson and Galbraith (1999) analyzed American wage data for the period from 1920 to 1947, allowing a direct test of the proposition (Goldin & Katz 2008) that improved education drove the "Great Compression" during the 1940s. We showed that practically all of the movement of relative wages across sectors in this period could instead be attributed to three identifiable forces: the movement of overall GDP in the Depression and war; the timing of labor actions, including especially strikes; and the movement of the exchange rate. This study reinforced the conclusions of *Created Unequal* (Galbraith 1998), which analyzed the evolution of weekly payrolls by industrial classification from 1958 into the 1990s, showing in general terms that movements in the relative position of major sectors affected differentially by a small number of major forces – macroeconomic and political forces – and in the composition of employment were the dominant influences behind a changing distribution of pay and earnings.

The argument over rising wage inequalities was at first mainly American; it was rooted in surveys of US household earnings. It had, however, an important corollary for European economists, who supposed that their own countries were not subject to fluctuations in relative hourly wage rates. Thus developed the "Euro-sclerosis" view of chronically high European unemployment, which held that this was because of relative wages that refused to adjust to the pressures of technology and trade. High-skilled workers in Europe were paid too little and low-skilled workers too much, and the result was a failure of European labor markets to clear at full employment. The fault was presumed to lie with national labor-market institutions, and thus with protections and rules enforced by national law.

Galbraith and Garcilazo (2004) show how to isolate the effect on European unemployment of conditions and events at three distinct levels. First, there is the local or regional level, to which local labor-market conditions, including the wage structure, are most relevant. Then there is the national level, which captures the influence of law and tradition in each European country. Last, there is the common influence of forces affecting employment at the continental level, whether these emanate from common European policies or from forces in the wider world. Country- and time--fixed effects thus capture the role of national institutions and of continental macroeconomic conditions. A review of the country-fixed effects is sufficient to dismiss the notion that major differences separate the major continental economies of Europe (including Scandinavia); measures of growth, wealth, population structure, and wage inequality are sufficient to account for differences in unemployment among these countries.

Meanwhile, a simple model at the local level takes up the question of theoretical interest: what is the effect on unemployment of wage inequality? Here, the orthodox theory should offer a plain prediction: regions with higher levels of inequality, other things being equal, should experience lower unemployment rates; their inequality measures should serve as *prima facie* evidence of flexibility. In fact, the results are the opposite: higher local inequality is associated, not strongly but very consistently, with higher rates of local unemployment.

Two theoretical perspectives cast light on the finding. One is the inequality-and-migration model of Harris and Todaro (1970), which pointed out that high wage differentials induce people to quit low-paid jobs (for example, in peasant agriculture) to seek the small number of better opportunities (for example, in urban factories). Since there is inevitably a surplus of applicants in this situation, unemployment must result.

The condition of migrant labor in modern China provides a vivid illustration of this process. Second, there is the “LO model” of Meidner and Rehn (1951), who argued that small egalitarian countries (such as Sweden) can force the pace of productivity gains by compressing their wage structures, effectively squeezing out low-productivity industries and inviting in those that can profitably pay the required wages. Ultimately, the higher wealth accumulated via this process makes it possible to employ a large part of the labor force in low-productivity sectors, such as public services or subsidized farming, or in training programs. In either model, relative wage equality promotes more higher rates of employment; both are quite consistent with the evidence.

As researchers in development economics grew frustrated with a search for meaningful Kuznets curves in the available collections of survey-based evidence of income and expenditure inequalities around the world, attention focused on an apparently related question: what is the relationship between inequality and economic growth? Here, two opposed positions appeared, each claiming support in the collections of survey data that became available in the 1990s. One model, associated with Forbes (2000), argued that higher inequalities produced higher subsequent rates of growth; with Victorian logic, the mechanism ran through a higher propensity to save of the wealthy segment of the population, and therefore a more rapid rate of capital accumulation. The alternative, advanced notably by Birdsall, Ross, and Sabot (1995) held that countries that reduced inequalities, say through universal education and land reforms, were and would be rewarded with higher growth rates. Here, the mechanism ran through the expected reward to human capital; more widely distributed returns were held to induce a more sustained productive effort from the working population.

Apart from the then-available data, which were sparse, noisy, and hard to interpret, a difficulty with both of these theories lies in the effort to relate a measure of a level – the degree of inequality – to a subsequent rate of change, namely the rate of growth. Were either of the inequality/growth theories correct, it should be possible for countries to raise their income levels indefinitely, relative to other countries, by keeping their inequality in the “correct” position. Yet we know this is not the case. The problem is similar to that of relating the position of the floor pedals to the speed of a car: even if one correctly distinguishes the accelerator from the brake, the car will not speed up indefinitely if the accelerator is held down.

Further, as a matter of logic, there cannot be at the same time a Kuznets--type relationship between the level of income and the level of inequality, and either a Forbes- or a Birdsall-style relationship between the level of inequality and the subsequent rate of growth. From any given starting point, either of the latter relationships will erase the former over time. So the discovery that a relationship of the Kuznets type actually exists in data relating to pay would lead, inexorably, to the rejection of a relationship between inequality levels and later rates of growth.

Galbraith and Berner (2001) and Galbraith (2012) present evidence that there is a decided Kuznets relationship between income levels and the inequality of pay – as measured in a data set based on the UNIDO Industrial Statistics. Most countries are on a downward-sloping Kuznets surface, with inequality that declines as incomes rise, and at rates closely associated with the overall rate of economic growth. For this reason, the rich countries (members of the OECD) have markedly lower inequality measures than are found in the developing world. There are countries on upward-sloping Kuznets surfaces, however. They include China, which remains in the throes of a vast shift from the countryside to the cities and is thus today the canonical case of the classic Kuznets argument. The evidence suggests that the United States also fits this pattern, for a different reason: as a supplier of advanced capital goods, scientific products, and financial services to the world economy, the US is in a position in which strong growth differentially favors those already at the top of the income ladder. Thus the Kuznets curve exists, having acquired an upward swing at the high end.

The existence of a dense, consistent body of evidence for the level and evolution of pay inequality since the early 1960s permits another useful inquiry, namely into the existence of a worldwide pattern of change in inequalities, which can be done by estimating the time coefficients in a fixed-effects model. Galbraith and Kum (2003) found that inequalities measured within countries showed no consistent worldwide trend until about 1971, following which they tended to decline until around 1980. There then followed a twenty-year period of massively increasing inequalities, peaking in 2000, followed by a modest decline. During the period of rising inequality, distinct regional patterns of intensity can be discerned: first in Latin America (and in Africa, where visible) in the early 1980s, then in Central Europe and the former Soviet Union, and finally, in the 1990s, in Asia, especially in China.

Both the turning points and the regional patterns strongly point to a straightforward interpretation: inequality was stabilized under Bretton Woods, fell in the worldwide commodities and debt boom of the 1970s, and then rose massively in the debt crises and the era of financial instability and speculative excess that followed. It peaked in 2000, and thereafter fell in some countries, especially those in Latin America that separated themselves from the Washington Consensus after the currency crises in Brazil in the 1990s and in Argentina in 2002; the pattern of declining inequality was first detected by Galbraith, Spagnolo, and Pinto (2007).

Recent work in this vein compares changes in industrial pay inequality with movements of exchange rates (Rossi & Galbraith 2016). A striking relationship emerges for many countries: exchange rate depreciation raises inequality in the structure of pay. The logic of this finding is entirely mechanical. All industries, in all countries, sell either predominantly to the internal market or predominantly to the outside world. In most cases, the average pay in export sectors is higher than it is for industries that compete with imports. A devaluation raises the local-currency revenue of the exporting sectors, while making no change in the revenues of the others. Those extra revenues are paid out (at least to a degree) within the sector. Therefore, as a mechanical matter, a devaluation increases the gap between high-earning exporters and lower-earning domestic-sales-oriented firms. This shows up clearly in the relationship between exchange rates and pay inequality, especially when the US dollar is used as the reference currency, and especially following trade liberalizations.

Further as Galbraith, Halbach, Malinowska, Shams and Zhang (2014) have shown, estimates of gross household income inequality derived from measures of industrial pay inequality succeed in closely tracking the available (but much less dense and consistent) survey measures of gross income inequality for a large spectrum of countries around the world. Therefore, we can establish a clear line of causal flow, which must run from exchange rates to inequalities in the structure of industrial pay, and thence to inequalities in the structure of gross household incomes. The reverse sequence is not plausible. We may conclude that a major factor driving the movement of inequality measures, for a broad spectrum of countries excluding only the largest developed nations and those subject to the rigidity of the Euro, is the effect of exchange rate movements as determined in international currency markets.

All in all, there is no support here for the analysis of pay inequality as a micro-based national labor-market or wage-adjustment phenomenon, for the notions that technology or education are fundamental drivers of rising inequality in pay, nor for the idea that flexibility in wage setting has any favorable bearing on employment. Nor does the evidence lend any support to the notion that raising (or lowering) inequality can act as a durable driver of economic growth. Instead, the evidence shows a global pattern to the rise in inequality, suggesting that common global forces are mainly responsible, and that they operate within same broad framework of intersectoral transitions (and changing intersectoral terms of trade) that Kuznets identified nearly sixty years ago. These are macroeconomic and financial forces in the short run, and over a longer horizon, they are the forces of structural change. The timing and composition of the changes observed within the last generation, and especially since 1980, point directly at the conduct of world financial governance, at the neoliberal counter-revolution in policy, at the setting of global interest rates, and at the incidence of debt crises and debt deflations, as the crucial worldwide forces at play. This picture is reinforced by investigations at the national level in widely dispersed countries. Rising inequality is a marker of credit booms, and therefore also a potent indicator of the danger of macroeconomic instability and crisis. The global policy implication is that the control of inequality and the control of unstable finance are substantially the same problem.

Bibliography

Baker, D., Glyn, A., Howell, D. and Schmitt, M. (2005) "Labor Market Institutions and Unemployment: A Critical Assessment of the Cross-country Evidence" in D. R. Howell (ed.) *Fighting Unemployment: The Limits of Free Market Orthodoxy*. New York: Oxford University Press.

Baldwin, P. (2011) *The Narcissism of Minor Differences: How America and Europe Are Alike*. New York: Oxford University Press.

Birdsall, N., Ross, D. and Sabot, R. (1995) "Inequality and Growth Reconsidered: Lessons from East Asia". *The World Bank Economic Review* 9 (3): 477–508, https://doi.org/10.1093/wber/9.3.477.

Bluestone, B. and Harrison, B. (1990) *The Great U-turn: Corporate Restructuring and the Polarizing of America*. New York: Basic Books.

Bound, J. and Johnson, G. (1992) "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations". *American Economic Review* 82 (3): 371–92.

DiNardo, J. and Pischke, J.-S. (1990) "The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too?". National Bureau of Economic Research Working Paper No. 5606. Cambridge, MA: National Bureau of Economic Research.

Ferguson, T. and Galbraith, James K. (1999) "The American Wage Structure: 1920–1947". *Research in Economic History* 19: 205–57.

Forbes, K. J. (2000) "A Reassessment of the Relationship between Inequality and Growth". *American Economic Review* 90 (4): 869–87.

Galbraith, James K. (1998) *Created Unequal: The Crisis in American Pay*. New York: Free Press.
Galbraith, James K. (2012) *Inequality and Instability: A Study of the World Economy Just Before the Great Crisis*. New York: Oxford University Press.
Galbraith, James K. and Berner, M. (eds) (2001) *Inequality and Industrial Change: A Global View*. New York: Cambridge University Press.
Galbraith, James K. and Garcilazo, E. (2004) "Unemployment, Inequality and the Policy of Europe: 1984–2000". *Banca Nazionale del Lavoro Quarterly Review* 57 (228): 3–28.
Galbraith, James K., Halbach, B., Malinowska, A., Shams, A. and Zhang, W. (2014) "UTIP Global Inequality Data Sets 1963–2008: Updates, Revisions and Quality Checks". UTIP Working Paper No. 68, http://utip.lbj.utexas.edu/papers/UTIP_68.pdf.
Galbraith, James K. and Kum, H. (2003) "Inequality and Economic Growth: A Global View Based on Measures of Pay". *CESifo Economic Studies* 49 (4): 527–56, https://doi.org/10.1093/cesifo/49.4.527.
Galbraith, James K., Spagnolo, L. and Pinto, S. (2007) "Economic Inequality and Political Power: A Comparative Analysis of Argentina and Brazil". *Business and Politics* 9 (1): 1–25.
Goldin, C. and Katz, L. F. (2008) *The Race between Technology and Education*. Cambridge, MA: Belknap Press of Harvard University Press.
Harris, J. R. and Todaro, M. P. (1970) "Migration, Unemployment and Development: A Two-Sector Analysis". *The American Economic Review* 60 (1): 126–42.
Kuznets, S. (1955) "Economic Growth and Income Inequality". *The American Economic Review* 45 (1): 1–28.
Meidner, R. and Rehn, G. (1951) *Fackföreningsrörelsen och den Fulla Sysselsättningen*. Stockholm.
Milanovic, B. (2007) *Worlds Apart: Measuring International and Global Inequality*. Princeton, NJ: Princeton University Press.
Rossi, D. and Galbraith, James K. (2016) "Exchange Rates and Industrial Wage Inequality in Open Economies". UTIP Working Paper No. 71, http://utip.lbj.utexas.edu/papers/UTIP_71.pdf.
Wood, A. J. B. (1994) *North-South Trade, Employment and Inequality*. Oxford: Clarendon Press.

Abstract

Globalizacja a nierówności – nowe spojrzenie

Artykuł jest syntezą poglądów na temat nierówności oraz rozwoju gospodarczego prezentowanych w literaturze, poczynając od Kuznetsa (1955), przez neoklasyczne modele dostosowania rynku pracy oraz modele uwzględniające zmiany technologiczne i edukacyjne z lat 90., po rezultaty najnowszych badań, w których wykorzystano kompleksowe, spójne zbiory danych o zasięgu globalnym. Są one istotne dla oceny zarówno podobnych, jak i odmiennych doświadczeń gospodarek krajów wysoko rozwiniętych, krajów transformujących gospodarkę i krajów rozwijających się w dobie globalizacji. Prace oparte na zbiorach danych opracowanych w ramach Programu Nierówności realizowanego na Uniwersytecie w Teksasie potwierdzają pogląd Kuznetsa na temat dużego znaczenia międzysektorowego przepływu dochodów, ale zawierają wiele odrębnych i nowych ustaleń. Ustalenia te są możliwe, ponieważ obecnie można śledzić przepływy regionalne i ponadnarodowe dochodów, badać wpływ globalnych czynników makroekonomicznych na nierówności dochodowe, a ostatnio także ocenić krytycznie rolę, jaką odgrywają kursy wymiany w ewolucji nierówności w otwartych gospodarkach.

Słowa kluczowe: globalizacja, nierówności, płace, rynki pracy, produktywność.

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 15 • 2016
ISSN 1642-168X
AOC, 2016; 15: 21–42
DOI: 10.15678/AOC.2016.1502

Katarzyna Kochaniak

THE NATURE AND DETERMINANTS OF HIGH-VALUE HOUSEHOLD DEPOSITS IN THE EURO AREA*

Abstract

This paper analyses the occurrence and nature of high-value deposits in selected euro area countries and the determinants of a household's propensity to possess them. The analysis is conducted on household-level data and based on logistic regression. Due to the rarity of high-value deposits in euro area households, the sample of households surveyed is balanced in accordance with G. S. Maddala's approach. The results reveal considerable diversification in high-value deposits as a proportion of total deposits in euro area households. Some of the features of households and reference persons, such as attitudes to financial risk, saving aims, overall investment preferences, and the priority accorded to deposits compared to other financial assets, suggest that high-value deposits are of a long-term nature. The study finds that wealth, and certain socio-demographic characteristics, have a statistically significant influence on the likelihood of holding high-value deposits.

Keywords: high-value household deposits, deposit outflows, credit institution funding, liquidity standards.
JEL Classification: G21, D14, G01.

1. Introduction

Following the recent financial crisis, greater importance has been attached to stable funding for credit institutions. The new approach has been reflected in European Union regulations developed since 2013.

Katarzyna Kochaniak, Cracow University of Economics, Department of Finance, Rakowicka 27, 31-510 Kraków, Poland, e-mail: katarzyna.kochaniak@uek.krakow.pl

* This paper presents the results of a research project conducted with the financial support of the National Science Centre in Poland (grant number: DEC-2013/11/D/HS4/04056). The analysis is based on data provided by the Eurosystem Household Finance and Consumption Survey.

The introduction of the Liquidity Coverage Ratio (LCR) and the Net Stable Funding Ratio (NSFR) provides a powerful indication of the importance of stable funding for credit institutions in both the short and long run. Funding must be sufficient to ensure that, even in periods of stress, these institutions perform safely. EU provisions (EU 2013a, 2013b) concentrate on retail deposits and on household deposits in particular. Not all deposits are, however, defined in the same way. The shocks experienced in 2008–09 revealed certain features of deposits that make them more vulnerable. One such feature is that deposits of at least EUR 500,000 are regarded as very high value deposits. Regardless of local resistance to shocks, local standards of living, or the financial investment preferences of local households, this threshold now applies throughout the EU. Does it make sense to apply uniform regulation to a group of heterogeneous countries?

Based on household-level data from the Eurosystem Household Finance and Consumption Survey (HFCS), the aim of this paper is to investigate high-value household deposits and their determinants in 15 euro area countries. Respondents' propensity to possess deposits of this kind are analysed using a logistic regression model with reference to Maddala's approach to balancing the sample in the case of rare events (Maddala 2006).

Attempts are made to find answers to the following questions:

1. What proportion of household deposits in the euro area countries are high-value deposits?
2. With respect to financial standing and socio-demographic features, do the owners of high-value deposits form a single group in the euro area?
3. What are the determinants of a household's propensity to possess high-value deposits in the euro area?

The following hypothesis is tested: Households with high-value deposits are associated with a particular financial standing and with specific socio-demographic features. Though the frequency of households with high-value deposits is not identical for each state in the euro area, it is still possible to identify common sets of characteristics which influence their propensity to possess high-value deposits. It may be the case that, while the EU's uniform regulatory framework makes domestic credit institutions more resilient, the benefits of the "one size fits all" approach are limited due to heterogeneity.

Following the introduction, the paper presents a survey of the related literature, before proceeding to an account of the regulatory approach to high-value household deposits. There follows a description of the data and methodology employed to study the occurrence of high-value household deposits in the euro area and the determinants of households' propensity to possess them. The results are then set out and conclusions are drawn.

2. Related Literature

The literature on the funding stability of credit institutions discusses matters of importance during short-term crises and over longer periods of time. They include sources of funding for banks (Diamond & Rajan 2001, Borio 2009, Huang & Ratnovski 2011), financial assets held by households (Du Caju 2013), the connections between the limits of deposit insurance systems (DIS) and the way that individuals perceive risk (Karas, Pyle & Schoors 2013, Brown, Guin & Morkoetter 2013, Acharya & Mora 2015), the relationship between deposit outflows and incidents of financial turmoil (Cussen, O'Leary & Smith 2012), the links between downturns on commercial paper markets and deposit transfers (Pennacchi 2006, Gatev, Schuermann & Strahan 2009), the impact of interest rates on deposit outflows (Acharya & Mora 2012) and the correlation between deposit outflows and loan availability (Acharya, Almeida & Campello 2013). Some papers compare countries according to the purposes of household saving, with a focus on deposits as a component of household financial asset portfolios (Teppa et al. 2015).

In view of the implementation of the EU's uniform post-crisis regulations, the question of the stability of household deposits has lost none of its relevance. The paper is among the first to investigate the uniform regulation from the perspective of high-value household deposits and their determinants.

3. The Regulatory Approach to High-Value Household Deposits

The EU regulatory framework on the funding stability of credit institutions was based on the Basel III Accord of December 2010 (BCBS 2010). For LCR, it distinguished stable deposits as those with low rates of outflows of 5% or 3% and less stable deposits as those with outflow rates of 10% (BCBS 2013). In both cases, the run-offs were assumed as minimum floors. The adoption of increased outflow proportions was left to individual jurisdictions, which would have a sharper picture of the behaviour of local depositors in a period of stress. Though the high-value deposits could be counted among the less stable, the Basel Committee on Banking Supervision did not indicate any particular threshold for them (BCBS 2013).

The framework of the EU regulations on the funding stability of credit institutions is available in the in the package “Capital Requirements Directive IV and Capital Requirements Regulation” issued in 2013. The quality of funding was discussed within the terms of LCR and NSFR.

In 2013, the package was supplemented by the European Banking Authority (EBA) guidelines on the assessment of LCR at the EU credit institutions (EBA 2013a). They listed the factors determining higher outflows of retail deposits, pointing out the significance of their value. If the sum of deposits held by one client at one entity was in excess of EUR 100,000, or above the limit of a local deposit guarantee scheme (and in any case no higher than EUR 500,000), the deposit was to be regarded as of high value. The EBA also proposed a category of very-high deposits exceeding EUR 500,000. The conclusion could be drawn that there was a high risk of outflows of high-value deposits and a very high risk of outflows of very high value deposits.

The detailed information regarding less stable retail deposits, including household deposits, in periods of stress was presented by the European Commission (EC) in its delegated act in 2014 (EU 2015). Deposits exceeding EUR 500,000 were defined as high-value deposits and presented as the ones liable to increased volatility. The additional category proposed by the EBA was therefore not adopted.

The evolution of the EU definition of deposits with increased outflows, and the lack of formal empirical analysis in the individual countries regarding this issue, raised doubts about the appropriateness of the adopted limit of EUR 500,000 and thereby prompted the present study.

4. Data and Methodology

Fifteen euro area states were studied: Austria, Belgium, Cyprus, Finland, France, Germany, Greece, Italy, Luxembourg, Malta, the Netherlands, Portugal, Slovakia, Slovenia and Spain. They were selected due to the availability of the required information. The study was based on quantitative and qualitative data from the Eurosystem HFCS and was focused on households possessing high-value deposits. It should be noted that the database did not provide information on whether the sums above EUR 500,000 were held in one or more credit institutions. The sets of variables were organised as follows:

1. Quantitative, describing the household's:

– size (N): number of household members (NHM), number of members in employment (NME), number of members aged 16+ (NM16+);

– wealth (W): gross income (WGI), net wealth[1] (WNW), high-value deposits (WHD), value of sight deposits (WSD), value of savings deposits (WVD), value of total real assets (WRA) such as real estate, vehicles and valuables; value of total financial assets excluding deposits (WFA);

2. Qualitative[2], describing the household's:

– investment attitude (IA): willing to take substantial financial risks and expecting to earn substantial returns (IA1), willing to take above average financial risks and expecting to earn above average returns (IA2), willing to take average financial risks to earn average returns (IA3), unwilling to take any financial risks (IA4);

– reasons for saving (S): purchase of own home (SPH), other major purchases (SOP) such as residences, vehicles and furniture; setting up a private business or financing investments in an existing business (SFB), investing in financial assets (SFA), providing for unexpected events (SUE), paying off debts (SPD), provision for old-age (SOA), education/support for children and grandchildren (SES), bequests (SBQ), taking advantage of state subsidies (SAS);

– wealth (W'), possession of: mutual funds (W'MF), publicly-traded shares (W'TS), bonds (W'BO), collateralised loans (W'CL), gifts or an inheritance (W'GI);

3. Quantitative, describing a reference person:

– age (A);

4. Qualitative, describing a reference person:

– the highest level of education completed (E): tertiary (ETR), upper-secondary (EUS), lower-secondary (ELS), primary or below (EPR);

– marital status (M): married (MAR), single/never married (MSI), consensual union on legal basis (MCU), widowed (MWI), divorced (MDI);

– labour status (L): doing regular work for pay/self-employed/working in family business (LSW), on sick leave, maternity leave or another type of leave (LSL), unemployed (LSU), student/pupil/unpaid intern (LSS), retired or in early retirement (LSR), permanently disabled (LSD), compulsory

[1] Net wealth is defined as the difference between total (gross) assets and total liabilities. Total assets consist of real assets and financial assets.

[2] The qualitative variable takes the value 1 or 0 to indicate the presence or absence of a categorical effect that can be expected to change the outcome.

military service or equivalent social service (LSM), fulfilling domestic tasks (LST), other: not working for pay (LSO);

– gender (G): male (GMA); female (GFE);

– wealth (W'), possession of: public pension plans (W'PP), a voluntary pension scheme (W'VP);

5. Qualitative, describing country of residence (C): Austria (AT), Belgium (BE), Cyprus (CY), Germany (DE), Spain (ES), Finland (FI), France (FR), Greece (GR), Italy (IT), Luxembourg (LU), Malta (MT), the Netherlands (NL), Portugal (PT), Slovenia (SI), Slovakia (SK).

The first part of the study elicited answers to the following questions: What is the position of large deposits among all household deposits in the euro area countries? Regarding their financial standing and socio-demographic features, do the owners of high-value deposits represent a single group in the euro area? A number of the above variables were used to identify the characteristics of households and reference persons with high-value deposits. The variables also had the potential to reveal whether the deposits were stable or unstable. The variables from the following sub-sets were applied: size (N): NHM, NME; wealth (W and W'): WRA, WFA, WHD, WSD, WVD, W'GI; declared attitudes to risk (IA): IA 1-4; saving aims (S): SPH, SOP, SFB, SFA, SUE, SPD, SOA, SES, SBQ, SAS; age of reference person (A).

The second part of the study attempted to answer the following question: What are the determinants of the propensity of euro area households to hold large deposits? A number of variables were employed to identify them. As well as numerical variables: wealth (W) and household size (N), these were dummies from the following subgroups: age (A), education (E), gender (G), declared attitudes to risk (IA), labour status (L), marital status (M), countries (C) and wealth (W'). The variables were applied in the logit model described by the following formula:

$$y_i^* = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + u_i,$$

where: y_i^* – latent variable; x_{ij} – explanatory variables ($i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, k$); β_j – regression parameters (β_0 – constant); u_i – random component.

All of the observations of high-value deposits were enabled to form a dummy Y, which represented the fact that households owned high-value deposits (if the household has a large deposit $Y = 1$, otherwise $Y = 0$). Hence, the variable y_i^* could be defined as a household's propensity to hold a deposit exceeding EUR 500,000 (Maddala 2014) or as the probability that a household possesses a high-value deposit resulting from the occurrence of particular characteristics (Ulman 2011).

The first wave of HFCS includes 265 households with large deposits. The significant difference between that number and the remaining households, which did not possess such deposits, prompted the decision to balance the sample (Maddala 2006). As a result, a random subset of a further 265 households, this time without large deposits, was taken. The final sample considered in the analysis therefore consisted of 530 households.

Both types of variable – quantitative and qualitative – were applied in this part of the study. Some of the quantitative variables were converted into categorical variables, that is: total real assets (WRA), gross income (WGI), net wealth (WNW), total financial assets excluding deposits (WFA) and age of reference person (A) (Podolec, Ulman & Wałęga 2008). As the levels of these features were highly diverse, they were divided into three categories: low, medium and high. The boundaries of the assignment of characteristics to a particular category were determined by the values of quantile 0.33 ($q_{0.33}$) and quantile 0.66 ($q_{0.66}$). The levels were defined as follows (Table 1): low level of the feature: $x < q_{0.33}$; medium level of the feature: $q_{0.33} \leq x \leq q_{0.66}$; high level of the feature: $x > q_{0.66}$.

Table 1. Numerical Characteristics of Selected Household Characteristics (in EUR)

	$q_{0.33}$	$q_{0.66}$
WRA	245,000	979,300
WGI	38,100	97,100
WNW	268,992	2,008,625
WFA	1,320	139,444

Source: author's own calculations based on Eurosystem HFCS data.

Table 1 shows that variables for 33% of the euro area households depicted values not exceeding the quotations for $q_{0.33}$, while the remaining 67% depicted at least these sums. The quantile $q_{0.66}$ means that the characteristics of 66% of the households represented the values up to the specified level and the remaining 33% of households represented at least these values. In the next step, the categorical variables were converted into dummies, which were applied in the logit model. They referred to the low and high levels of the characteristics. The medium level was adopted as the base for comparison. As a consequence, total real assets (WRA) were converted into:

WRA LOW with a value of 1 when WRA < 245,000 and 0 in all other cases; WRA MEDIUM with a value of 1 when 245,000 ≤ WRA ≤ 979,300 and 0 in all other cases; WRA HIGH with a value of 1 when WRA > 979,300 and 0 in all other cases. The remaining variables, except age (A), were treated in the same way. The quantiles for age indicated only the boundaries deciding the assignment of households to particular categories. The variable A LOW took a value of 1 when A ≤ 50 years old and 0 in all other cases; A MEDIUM took a value of 1 when A ≤ 65 years old and 0 in all other cases; A HIGH took a value of 1 when A > 65 years old and 0 in all other cases.

Parameter estimates from multiple regression models[3] were used as initial values of the parameters in the logit models.

5. Results

The first part of the study analysed the significance of high-value deposits (WHD) and the characteristics of the depositors in individual countries.

The data disclosed that only small proportions of households declared high-value deposits in the national samples which, in the Slovakian and Slovenian cases, revealed no high-value deposits whatsoever. The proportion of households holding high-value deposits was greatest in the Spanish case (2%). These proportions were consistent with the share of all high-value deposits in total deposits at the domestic level (Figure 1). In some member states, however, the impact of high-value deposits on the overall total was significant. In Spain, for example, high-value deposits constituted 41.3% of total household deposits, while they made up 25.9% of total household deposits in Belgium and 22.7% in Luxembourg. The proportion of high--value deposits in total deposits was, at 10–15%, also notable in Cyprus, France, Austria and Portugal. These results indicate that high-value deposits play an important role in the funding of selected MFI sectors in the euro area. Building detailed profiles of the people who hold them could be the key to understanding the volatility of high-value deposits in the individual euro area states.

There was considerable diversification in the amounts of high-value deposits held in the euro area (Table 2), which was expressed in a coefficient of variation (CV) of 12–124%. The highest amounts were noted in Belgium and Luxembourg and the lowest in Cyprus and Finland. The median for high-value deposits varied from EUR 550,000 in Cyprus to EUR 833,357 in Spain. The minimum levels were close to the adopted threshold almost

[3] The explanatory variables were selected based on stepwise regression.

everywhere, while the maximum levels lay in a range from EUR 700,000 in Cyprus and EUR 7,050,000 in Spain.

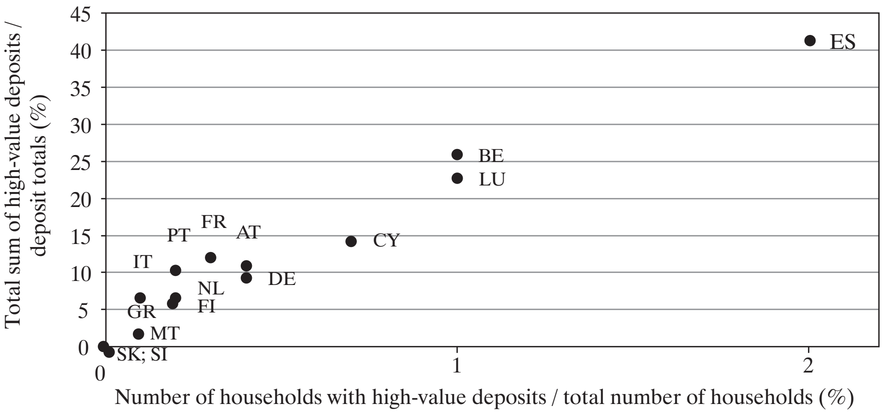

Fig. 1. Sum of High-value Household Deposits/Household Deposit Totals (%) and Number of Households with High-value Deposits/Total Number of Households (%) in Individual Eurozone Countries

Source: author's own study based on Eurosystem HFCS data.

It can be assumed that at least some proportion of the diversity between countries was the result of the heterogeneity of households, which was described by characteristics such as size (NHM, NME), wealth (WRA, WFA, WSD, WVD, W'GI), declared attitudes to risk (IA), saving aims (S) or age of the reference person (A). The variables are summarised in Table 3 and Table 4.

The average number of household members did not exceed three in any of the countries studied. In only a few of them, however, was there more than one person in employment. The most frequent attitudes to risk were "willing to take average financial risks" and "unwilling to take any financial risks". These outlooks could suggest that high-value deposits are stable. If that were so, the regulatory approach may not be appropriate. The most common saving aims of the households surveyed – provision for old--age and bequests – also revealed the long-term nature of the deposits. The households demonstrated considerable diversification in the total real assets and total financial assets they held. It should be noted that past gifts and inheritances accounted for the financial position of a large proportion of households, which was therefore not the result of wise investment decisions. Deposits dominated financial assets in Belgium, Finland, France, Italy,

Table 2. Summary Statistics of High-value Deposits in Households in Selected Eurozone Countries

	Number of observations	Household deposits						
		WHD					WSD	WVD
		mean*	median*	min*	max*	cv**	mean*	mean*
AT	9	791,853	630,200	527,000	1,444,834	45	49,957	741,896
BE	25	1,162,651	620,800	500,000	5,003,000	107	132,085	1,030,566
CY	9	561,144	550,000	500,000	700,000	12	72,255	488,889
DE	13	846,154	603,000	500,001	2,550,000	69	86,615	759,538
ES	118	1,066,683	833,357	500,000	7,050,000	80	275,905	790,778
FI	18	670,661	599,159	588,409	820,831	16	670,661	****
FR	51	924,385	762,114	500,600	2,683,490	45	577,374	347,010
GR	1	1,300,000	***	***	***	***	1,300,000	0
IT	10	723,757	607,187	501,484	1,226,905	37	409,161	314,596
LU	12	1,093,965	606,809	515,000	5,286,768	124	119,252	974,713
MT	1	515,142	***	***	***	***	15,141	500,001
NL	2	1,731,256	***	597,012	2,865,500	***	5,251	1,726,005
PT	7	1,065,571	787,500	516,000	1,935,000	57	77,571	988,000
EA	276	979,816	700,247	500,000	7,050,000	80	315,196	710,989

Note: cv (coefficient of variation) – standard deviation/arithmetic mean. * In EUR. ** In %. *** Number of responding households was too small to estimate statistics. **** No information available in the database.

Source: author's own calculations based on Eurosystem HFCS data.

Table 3. Summary Statistics of Selected Household Characteristics in Eurozone Countries

	Number of observations	Household characteristics										
		N (mean)			IA		WGI*				S	
		NHW	NME	NM16+	IA3 & IA4**	GI yes**	mean	median	min	max	SOA yes**	SBQ**
AT	9	3	1	3	33	78	215,177	100,500	53,401	515,762	78	22
BE	25	2	0	2	76	32	87,005	76,320	18,100	310,000	48	36
CY	9	3	2	3	89	67	151,734	91,000	55,050	432,000	56	22
DE	13	2	1	2	92	62	432,088	336,000	49,000	1,350,000	54	0
ES	118	2	1	2	92	67	375,621	108,349	0	8,760,318	19	11
FI	18	2	2	2	****	****	265,666	12,683	85,425	1,080,636	****	****
FR	51	2	0	2	****	69	236,388	125,848	18,822	1,833,944	****	****
GR	1	3	2	2	0	0	2,110,000*	***	****	***	0	0
IT	10	2	1	2	70	****	87,234	79,068	9,531	194,300	****	****
LU	12	3	1	3	100	50	252,234	250,100	31,200	520,560	42	42
MT	1	2	0	2	100	0	51,000*	***	***	***	0	100
NL	2	2	1	2	100	0	185,861*	***	128,734	242,988	100	100
PT	7	3	2	2	86	29	189,660	125,000	38,930	604,600	33	33
EA	276	2	1	2	64	55	289,909	122,741	0	8,760,318	23	13

* In EUR. ** In %. *** Number of responding households was too small to estimate statistics. **** No information available in the database.

Source: author's own calculations based on Eurosystem HFCS data.

Table 4. Summary Statistics of Household Wealth in Selected Eurozone Countries

	Number of observations	W							
		WFA*				WRA*			
		mean	median	min	max	mean	median	min	max
AT	9	983,046	944,798	617,784	1,448,877	1,404,053	870,000	499,747	3,412,000
BE	25	684,117	400,750	0	2,100,000	1,266,706	574,500	0	6,402,814
CY	9	565,327	139,444	500	3,100,000	7,579,859	6,492,000	2,674,169	13,065,000
DE	13	2,812,856	1,040,000	600,000	13,555,000	6,402,577	1,804,000	685,000	62,750,000
ES	118	1,656,513	253,389	0	33,000,000	7,065,044	1,964,102	128,400	94,003,000
FI	18	265,376	251,591	7,503	651,102	1,620,517	721,795	148,828	11,225,655
FR	51	703,298	275,558	0	4,427,245	1,701,840	1,175,115	33,593	16,397,098
GR	1	300,000	**	**	**	10,100,000	**	**	**
IT	10	288,503	22,500	0	1,762,784	1,098,864	572,000	72,000	5,540,000
LU	12	375,582	150,205	0	1,580,133	2,190,750	1,460,000	453,500	5,877,000
MT	1	872,058	**	**	**	25,000	**	**	**
NL	2	665,613	**	1,010,780	320,446	377,258	**	225,000	529,516
PT	7	550,679	143,750	0	2,498,000	5,796,214	1,900,000	210,300	210,300
EA	276	1,084,573	255,764	0	33,000,000	4,483,733	1,424,852	0	94,003,000

* In EUR. ** Number of responding households was too small to estimate statistics.

Source: author's own calculations based on Eurosystem HFCS data.

Luxembourg, the Netherlands and Portugal. Other than in the Netherlands and Malta, the average value of real assets exceeded the average value of financial assets (excluding deposits). This may suggest that lower-risk investment in real estate, vehicles or valuables – rather than in riskier mutual funds, bonds or publicly-traded shares – was the priority of households holding high-value deposits. The low-volatility of high-value deposits over the long term would appear to be confirmed by the average age of the reference persons and the reasons for saving that were most often given. The age of the reference persons was 49–66 in almost all of the countries[4] and, following modifications adopted in recent years, was lower than the retirement age.

The first part of the study demonstrated that the role of high-value deposits in the funding of credit institutions was based on individuals at the domestic level. Furthermore, some of the characteristics of households suggested that this role was relatively stable. Insofar as it may prevent credit institutions from treating the deposits as stable and compel a report of reduced stability of funding to be submitted, the implementation of a uniform threshold of EUR 500,000 may become a burden in countries where high-value deposits account for the greatest proportion of total deposits.

The priority of the second part of the study was to identify sets of common features influencing the probability that a euro area household will possess a high-value deposit. These features can also be understood as ones that affect the propensity to hold a large deposit (Ulman 2011).

The insufficient number of households with high-value deposits in some of the national samples forced the removal from the study of countries with less than nine such cases. A group of nine euro area countries, made up of Austria, Belgium, Cyprus, Germany, Spain, Finland, France, Italy and Luxembourg, remained to be analysed.

The study implemented four versions of the logit model, in which the potential factors describing the probability that a euro area household will possess a high-value deposit were:

1. WNW LOW, WNW HIGH – highlighting the significance of net wealth;

2. WNW LOW, WNW HIGH, WGI LOW, WGI HIGH, A LOW, AT, BE, CY, ES, FI, FR, IT, LU – taking account of net wealth and gross annual

[4] AT – 51, BE – 65, CY – 58, DE – 64, ES – 66, FI – 59, FR – 71, GR – 41, IT – 60, LU – 61, MT – no data available, NL – 71, PT – 63.

income (the main driver of net wealth), the lowest and highest values for the age for reference persons, and country affiliation;

3. WRA LOW, WRA HIGH, WFA LOW, WFA HIGH, A LOW, A HIGH, AT, BE, CY, ES, FI, FR, IT, LU – this is a modification of version 2 that refers to age, country of residence[5], and to real and financial assets instead of net wealth and gross income;

4. MEM 16+, ETR, MAR, GMA, A LOW, A HIGH, AT, BE, CY, ES, FI, FR, IT, LU – referring to the socio-demographic features of a household and the country that it is in[6].

The first version of the logit model was the simplest. It tested the influence of household net wealth on the probability that a household would possess a high-value deposit. The WNW HIGH appeared as a statistically significant explanatory variable (Table 5). A household's propensity to possess a high-value deposit increased when it appeared in the highest range for net wealth. This indicates that a way of living that places a high value on asset collection, while being wary of consumption and debt, was the favoured tendency. It can also be concluded that high-value deposits were an attribute of the most affluent households.

Table 5. Summary of Logistic Regression Analysis for Variables Predicting the Incidence of High-value Deposits in Households (Version 1)

	B	*SE B*	*t*(517)	*p*-value
Constant	–1.0312	0.1214	–8.4917	0.0000
WNW HIGH	4.2386	0.4043	10.4848	0.0000
Odds ratio = 69.31; correctly classified households – 81.32%; chi-square (11) = 272.34; $p < 0.0000$				

Source: author's own calculations based on Eurosystem HFCS data.

The odds ratio[7] confirmed that the classification of households in the analysed category was more precise than a random selection (the probability of correctly classifying households by this model was 69.31 times higher than of incorrectly classifying them). More than 80% of households were correctly classified. The likelihood ratio (LR) Chi-Square test confirmed the significant influence on the propensity to possess a high value deposit of

[5] Germany was the base for comparison.

[6] See footnote 5.

[7] The odds ratio is defined as a multiplication of correctly classified observations in relation to a multiplication of incorrectly classified ones, with a given vector x_i of explanatory variables.

the variable under consideration and thus rejected the hypothesis that such an effect was absent.

The second version of the logit model, which employed a set of potential explanatory variables that included net wealth, gross income (the primary driver of net wealth), age of reference person and country of residence, confirmed the conclusions of the first. It was found that the propensity to possess a high-value deposit was weakest among households with low gross income and low net wealth. Whereas the probability of holding a high-value deposit decreased significantly where the reference persons were young, it increased where the factors of high net wealth and high gross income were present. Assuming that other independent variables hold constant, high-value deposits were more frequent in Spanish, Belgian and Austrian households than in those of the remaining countries analysed. It can be stated that the wealthiest households, that is, those in Spain, Belgium and Austria, whose reference persons were 51 years old, had the strongest propensity to possess high-value deposits. The results for version 2 of the logit model are presented in Table 6.

Table 6. Summary of Logistic Regression Analysis for Variables Predicting the Incidence of High-value Deposits in Households (Version 2)

	B	*SE B*	*t*(517)	*p*-value
Constant	–0.0767	0.2754	–0.2787	0.7806
WNW HIGH	2.4103	0.4481	5.3788	0.0000
WNW LOW	–27.6137	5530.3410	–0.0050	0.9960
WGI HIGH	0.8041	0.3853	2.0869	0.0374
WGI LOW	–1.3380	0.4250	–3.1485	0.0017
A LOW	–1.3665	0.3963	–3.4483	0.0006
ES	1.5464	0.4118	3.7554	0.0002
BE	2.1583	0.6625	3.2576	0.0012
AT	3.0382	1.2213	2.4877	0.0132
Odds ratio = 61.63; correctly classified households – 88.68%; chi-square (8) = 482.67; $p < 0.0000$				

Source: author's own calculations based on Eurosystem HFCS data.

In this version, the odds ratio confirmed that the classification of households was better than a random classification with regard to the presented categories. More than 90% of households were correctly classified. The likelihood ratio (LR) Chi-Square test confirmed that the set of variables under consideration had a significant influence on the propensity of households to hold high-value deposits. The hypothesis of the absence of such effects was thus rejected.

It was found in the third version of the logit model that not all of the potential explanatory variables (components of net wealth: real and financial assets excluding deposits) entered the model. Only WRA HIGH, WRA LOW, WFA HIGH, A LOW, ES, BE, and AT were statistically significant. The results are set out in Table 7. It should be noted that the dummies indicated that the same countries of residence as in version 2 had a positive impact on propensity to hold a high-value deposit. This means that the probability of possessing a high-value deposit appeared to be greater in those countries than in the remaining member states. The same conclusion could be drawn from the WRA HIGH and WFA HIGH variables. The propensity to possess a high-value deposit increased when a household had high real and financial assets. Conversely, if households did not tend to accumulate real assets, the probability of holding a high-value deposit decreased. Explanatory variable A LOW demonstrated that the willingness to possess high-value deposits among young reference persons was lower than in the other groups. This version of the logit model told us that real and financial assets in the highest range of classification, reference persons aged over 50 and residence in Spain, Belgium or Austria were the variables most strongly related in the euro area to holding high-value deposits. The results are set out in Table 7.

Table 7. Summary of Logistic Regression Analysis for Variables Predicting the Incidence of High-value Deposits in Households (Version 3)

	B	*SE B*	*t*(517)	*p*-value
Constant	–0.8077	0.2463	–3.2798	0.0011
WRA HIGH	1.9047	0.3499	5.4434	0.0000
WRA LOW	–1.5571	0.3432	–4.5367	0.0000
WFA HIGH	1.9627	0.3263	6.0149	0.0000
A LOW	–1.4526	0.3345	–4.3426	0.0000
ES	1.2556	0.3292	3.8140	0.0002
BE	1.8107	0.5309	3.4104	0.0007
AT	1.5036	0.6419	2.3425	0.0195

Odds ratio = 49.39; correctly classified households – 87.52%; chi-square (7) = 379.25; $p < 0.0000$

Source: author's own calculations based on Eurosystem HFCS data.

As in previous versions, the odds ratio informed us that better results were obtained from households that had been classified rather than randomly classified. The probability of carrying out a correct classification of households based on this model was 49 times higher than of carrying out an incorrect one. Almost 90% of households were classified correctly. The likelihood ratio (LR) Chi-Square test confirmed that the set of variables under consideration had a significant influence on the propensity of households to hold high-value deposits and rejected the hypothesis that such effects were absent. It can therefore be concluded that all of the household characteristics implied in the model – real and financial assets in the highest range of classification, reference persons aged over 50 and residence in Spain, Belgium or Austria – had a significant impact on the propensity to possess high-value deposits in the area analysed.

The fourth version concerned only the impact of socio-demographic features on the willingness of households to hold high-value deposits. The following potential explanatory variables were statistically significant: NM16+, ETR, LSR, GMA, A LOW, A HIGH, ES, AT, BE, LU, CY, and FR. It should be noted that the dummies for these countries of residence had a positive impact on the probability that a household would possess a high-value deposit[8]. This means that the likelihood of possessing a high-value deposit was greater than in Germany, Finland and Italy. Because there was a greater chance that more of its members would be in employment, the probability of possessing a high-value deposit was boosted when the number of household members aged 16 or over was greater. The propensity to possess a high-value deposit also increased if the gender of the household's reference person was male and that person was at least 65 years old. Where the reference persons were young, though, the probability of holding a high-value deposit was lower. The probability was higher where reference persons had completed tertiary education. The propensity to hold high-value deposits among retired reference persons was lower than for groups belonging to other employment classifications. This version of the logit model told us that households with a greater number of members aged over 16 who are well-educated, male, aged over 65 and still in employment were most likely to hold high-value deposits. The results of the analysis are presented in Table 8.

The odds ratio confirmed that the classification of households into the listed categories yielded better results than random classification.

[8] In relation to the countries that formed the base for comparison.

The probability of carrying out a correct classification of households by this model was 16 times greater than of performing an incorrect classification. Eighty per cent of households were classified correctly. The likelihood ratio (LR) Chi-Square test confirmed that the set of variables under consideration had a significant influence on the propensity of households to hold high--value deposits and rejected the hypothesis that such effects were absent. In conclusion, the socio-demographic features of households examined in this version of the model had a significant influence on the willingness of euro area households to hold high-value deposits.

Table 8. Summary of Logistic Regression Analysis for Variables Predicting the Incidence of High-value Deposits in Households (Version 4)

	B	*SE B*	*t*(517)	*p*-value
Constant	–2.7030	0.4882	–5.5377	0.0000
ES	1.9893	0.3242	6.1346	0.0000
AT	2.3697	0.6077	0.8997	0.0001
BE	1.7913	0.4772	3.7534	0.0002
CY	1.4646	0.3345	1.9873	0.0473
FR	0.6655	0.3137	2.1215	0.0343
LU	2.1608	0.6735	3.2084	0.0014
NM16+	0.5087	0.1554	3.2719	0.0011
LSR	–0.9675	0.3664	–2.6401	0.0085
ETR	1.6854	0.2493	6.7602	0.0000
GMA	0.6954	0.2469	2.8121	0.0051
A LOW	–2.2266	0.3507	–6.3495	0.0000
A HIGH	0.9877	0.3634	2.7179	0.0068
Odds ratio = 16.00; correctly classified households – 80.00%; chi-square (12) = 272.41; $p < 0.0000$				

Source: author's own calculations based on Eurosystem HFCS data.

Though all of the versions of the logit model identified characteristics of euro area households likely to possess high-value deposits, it was difficult to point to the one with the best fit in the statistical sense: they all described the problem very well and from different perspectives. The results confirm the major impact of household wealth on the probability of holding a high--value deposit. Versions two, three and four detected that the youngest age category had a negative impact, and the two older categories a positive impact, on the propensity to hold high-value deposits. Of the countries under consideration, it was found that the households of Belgium, Austria and Spain had a significantly stronger propensity to possess high-value deposits. Version four of the logit model told us that households with the bracketed socio-demographic characteristics (a greater number of members aged over 16 who are well-educated, male, aged over 65 and still in employment) were most likely to hold high-value deposits.

6. Conclusions

The last financial crisis revealed the importance of stable funding if credit institutions are to be resilient and able to dispose of liquidity shocks. The current EU regulations describe household deposits of up to EUR 500,000 as stable. As they may impose unnecessary limits in countries where funding is reported as stable by supervisory authorities, such precise guidelines may, however, prove ineffective. The uniform threshold may thus become merely a needless benchmark serving only to make the regulation more complicated.

The first part of the study identified considerable diversification of high-value deposits in the households of the group of countries analysed. The proportion of respondents who declared that they held them was small: no greater than 2%. In some member states, however, the impact of high-value deposits on the overall total was significant. In Spain, for example, high-value deposits constituted 41.3% of total household deposits, while they made up 25.9% of total household deposits in Belgium and 22.7% in Luxembourg. The scale of these shares in funding did not imply, though, that household deposits were of poor quality. Particular features of households and reference persons, such as attitudes to financial risk, saving aims, a focus on real rather than financial assets, the prioritising of deposits among financial assets and the age of the reference persons, suggested instead that high-value deposits were of a long-term nature.

The second part of the study, which presented certain sets of features thought likely to increase the propensity of euro area households to possess high-value deposits, clearly demonstrated the significance of household wealth and socio-demographic characteristics for the occurrence of this propensity. Net wealth and its components appeared as features of primary importance, which leads to the conclusion that a style of life that places great value on accumulating financial and real assets was responsible for the availability of large deposits for credit institutions. The tendency to hold high-value deposits was relatively low among households with young reference persons. However, the need to possess large deposits in the other age subgroups may have been caused by the decreasing incomes that characterise later life. The logit model informed us that households in Belgium, Spain and Austria had a greater propensity, or capacity, to accumulate high-value deposits. It should be noted that the euro area states identified in the first part of the study were those whose credit institutions held the greatest amount of high-value deposits as a proportion of total household deposits. It is assumed that as a consequence the EUR 500,000 threshold may have a more negative influence on the funding stability of Belgian, Spanish and Austrian entities than it does on credit institutions in the remaining states.

Bibliography

Acharya, V., Almeida, H. and Campello, M. (2013) "Aggregate Risk and the Choice between Cash and Lines of Credit". *The Journal of Finance* 68 (5): 2059–2116, https://doi.org/10.1111/jofi.12056.

Acharya, V. and Mora, N. (2012) "Are Banks Passive Liquidity Backstops? Deposit Rates and Flows During the 2007–2009 Crisis". NBER Working Paper No. 17838, http://www.nber.org/papers/w17838.

Acharya, V. and Mora, N. (2015) "A Crisis of Banks as Liquidity Providers". *The Journal of Finance* 70 (1): 1–43, https://doi.org/10.1111/jofi.12182.

BCBS (2010) *Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems.* Basel: Bank for International Settlements.

BCBS (2013) *Basel III: The Liquidity Coverage Ratio and Liquidity Risk Monitoring Tools*. Basel: Bank for International Settlements.

Borio, C. (2009) "Ten Propositions about Liquidity Crises". BIS Working Paper No. 293. Basel: Bank for International Settlements, pp. 1–21.

Brown, M., Guin, B. and Morkoetter, S. (2013) "Deposit Withdrawals from Distressed Commercial Banks". Swiss Institute of Banking and Finance Working Papers on Finance No. 19. St. Gallen: Swiss Institute of Banking and Finance, pp. 1–33.

Cussen, M., O'Leary, B. and Smith, D. (2012) "The Impact of the Financial Turmoil on Households: A Cross Country Comparison". *Central Bank of Ireland Quarterly Bulletin* 2 (April): 78–98.

Diamond, D. W. and Rajan, R. G. (2001) "Banks and Liquidity". *American Economic Review* 91 (2): 422–25.

Du Caju, P. (2013) "Structure and Distribution of Household Wealth: An Analysis Based on the HFCS". *National Bank of Belgium Economic Review* 9: 41–62.

EBA (2013a) "Guidelines on Retail Deposits Subject to Different Outflows for Purposes of Liquidity Reporting under Regulation (EU) No 575/2013 on Prudential Requirements for Credit Institutions and Investment Firms and Amending Regulation (EU) No 648/2012 (Capital Requirements Regulation – CRR)". London: European Banking Authority.

EBA (2013b) "EBA Discussion Papers on Defining Liquid Assets in the LCR under the Draft CRR and on Retail Deposits Subject to Higher Outflows for the Purposes of Liquidity Reporting". London: European Banking Authority.

ECB (2013) "The Eurosystem Household Finance and Consumption Survey – Results from the First Wave". *ECB Statistics Paper Series* 2 (April): 1–112.

EU (1994) "Directive 94/19/EC of the European Parliament and of the Council of 30 May 1994 on Deposit-Guarantee Schemes", OJ L 135, 31 May 1994.

EU (2013a) "Directive 2013/36/EU of the European Parliament and of the Council of 26 June 2013 on Access to the Activity of Credit Institutions and the Prudential Supervision of Credit Institutions and Investment Firms, Amending Directive 2002/87/EC and Repealing Directives 2006/48/EC and 2006/49/EC", OJ L 176, 27 June 2013.

EU (2013b) "Regulation (EU) No 575/2013 of the European Parliament and of the Council of 26 June 2013 on Prudential Requirements on Credit Institutions and Investment Firms and Amending Regulation (EU) No 648/2012, OJ L 176, 27 June 2013.

EU (2015) "Commission Delegated Regulation (EU) 2015/61 of 10 October 2014 to Supplement Regulation (EU) No 575/2013 of the European Parliament and the Council with Regard to Liquidity Coverage Requirement for Credit Institutions", OJ L 11, 17 January 2015.

EU (2014) "Directive 2014/49/EU of 16 April 2014 on Deposit-Guarantee Schemes", OJ L 173, 12 June 2014.

Gatev, E., Schuermann, T. and Strahan, P. (2009) "Managing Bank Liquidity Risk: How Deposit-Loan Synergies Vary with Market Conditions". *Review of Financial Studies* 22 (3): 995–1020, https://doi.org/10.1093/rfs/hhm060.

Huang, R. and Ratnovski, L. (2011) "The Dark Side of Bank Wholesale Funding". *Journal of Financial Intermediation* 20 (2): 248–63, https://doi.org/10.1016/j.jfi.2010.06.003.

Karas, A., Pyle, W. and Schoors, K. (2013) "Deposit Insurance, Banking Crises, and Market Discipline: Evidence from a Natural Experiment on Deposit Flows and Rates". *Journal of Money*, *Credit and Banking* 45 (1): 179–200, https://doi.org/10.1111/j.1538-4616.2012.00566.x.

Maddala, G. S. (2006) *Ekonometria*. Warszawa: Wydawnictwo Naukowe PWN.

Mody, A. and Sandri, D. (2011) "The Eurozone Crisis: How Banks and Sovereigns Came to be Joined at the Hip". IMF Working Paper No. 11 (269), Washington, DC: International Monetary Fund, pp. 1–33.

Pennacchi, G. (2006) "Deposit Insurance, Bank Regulation, and Financial System Risks". *Journal of Monetary Economics* 53 (1): 1–30, https://doi.org/10.1016/j.jmoneco.2005.10.007.

Podolec, B., Ulman, P. and Wałęga, A. (2008) *Aktywność ekonomiczna a sytuacja materialna gospodarstw domowych* [Economic activity and the material situation of households]. Kraków: Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie.

Teppa, F., Porpiglia, A., Ziegelmeyer, M., Le Blanc, J. and Zhu, J. (2015) "Household Saving Behaviour and Credit Constraints in the Euro Area". ECB Working Paper No. 1790. Luxembourg: Publications Office, pp. 1–62.

Ulman, P. (2011) *Sytuacja ekonomiczna osób niepełnosprawnych i ich gospodarstw domowych w Polsce* [The economic situation of disabled people and their households in Poland]. Zeszyty Naukowe – Seria specjalna: Monografie nr 199. Kraków: Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie.

Abstract

Jakość dużych depozytów detalicznych i determinanty ich występowania w krajach strefy euro

W artykule zaprezentowano jedną z regulacyjnych kategorii depozytów detalicznych – tzw. duże depozyty gospodarstw domowych i podjęto próbę oceny ich wrażliwości na odpływ. W pracy opisano także czynniki determinujące skłonność gospodarstw domowych do posiadania takich aktywów.

Analizę przeprowadzono na podstawie danych jednostkowych o gospodarstwach domowych pochodzących z bazy Eurosystemu Household Finance and Consumption Survey. W celu identyfikacji determinant występowania dużych depozytów zastosowano metodę regresji logistycznej. Ze względu na fakt, że depozyty należące do opisywanej kategorii stanowią rzadką cechę gospodarstw domowych, badanie przeprowadzono na próbie zbilansowanej, zgodnie z podejściem proponowanym przez G. S. Maddalę.

Uzyskane wyniki wskazują na zróżnicowane udziały dużych depozytów w ogóle depozytów deklarowanych przez gospodarstwa domowe w poszczególnych krajach strefy euro. Ponadto pozwoliły one określić cechy gospodarstw domowych odnoszące się do sytuacji finansowej oraz cech społeczno-demograficznych, które istotnie wpływają na prawdopodobieństwo posiadania dużych depozytów.

Słowa kluczowe: depozyty gospodarstw domowych, duże depozyty, normy płynności, źródła finansowania banków.

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 15 • 2016
ISSN 1642-168X
AOC, 2016; 15: 43–56
DOI: 10.15678/AOC.2016.1503

Beata Ciałowicz

Andrzej Malawski

THE LOGIC OF IMITATIVE PROCESSES: IMITATION AS SECONDARY INNOVATION – AN AXIOMATIC SCHUMPETERIAN ANALYSIS

Abstract

This paper offers an axiomatic analysis of the imitative activity of producers in a Schumpeterian process of innovative change. It argues that structural change in the economy is generated by leaders of radical innovation, whose actions trigger the diffusion of innovations, and whose strategies and innovations can be copied by imitators. As a consequence, these imitators become second-order innovators operating in a production system that is deprived of primary innovators. The paper demonstrates that increases in the number and variety of second-order innovators can intensify innovative changes throughout the production system. Furthermore, this logic can be reconstructed by reference to the research programme on modelling Schumpeterian innovative evolution within the Arrow-Debreu dynamic general equilibrium theory.

Keywords: imitation, innovation, production system, axiomatic analysis, Schumpeterian approach.
JEL Classification: O31, O10, C6.

1. Introduction

The concept of imitation would appear to be ubiquitous in current evolutionary studies (Bessen & Maskin 2009, Glass 2010, Mukoyama 2003, Segestrom 1990, Shenkar 2010). Herrmann-Pillath (2013) states that at the

Beata Ciałowicz, Cracow University of Economics, Faculty of Finance and Law, Department of Mathematics, Rakowicka 27, 31-510 Kraków, Poland, e-mail: beata.cialowicz@uek.krakow.pl

Andrzej Malawski, Cracow University of Economics, Faculty of Finance and Law, Department of Mathematics, Poland

fundamental, phylogenetic level the role of imitation in the co-evolution of human brain size and human group size is crucial. Imitation, which is the basic form of human learning during ontogeny, is a uniquely human capacity. In contrast to apes, writes the German sinologist: "Human infants develop the capacity to imitate others' actions in a context-free way, that is, they become able to separate goals of actions from individual perspectives, and they can replicate intentions in their own actions" (2013, pp. 225–26). He also indicates the methodological significance of imitation which, if simply conceived as copying, allows us to model imitative copies by using techniques applied earlier to their originals.

Nelson and Winter (1982) claim that imitation is an important mechanism underlying the behaviour of firms. According to Safarzyńska and van den Bergh, it makes possible: "(…) savings on the costs of individual learning, experimentation or searching by exploiting information already acquired by others" (2010, p. 351). In this reading, copying can either mean replicating the most successful strategy or duplicating the strategy adopted by the majority.

By way of contrast, discussion of the Schumpeterian triad of invention, innovation and imitation has been dominated by the view that only its first two elements play a significant role in economic development. The part played by imitation has thus been somewhat neglected (Andersen 2009, Hanusch & Pyka 2006, 2007). Niosi (2012), however, suggests that the idea is ubiquitous in the evolutionary dynamics of industry. To explain its role in catching-up processes, for example, he discusses imitative innovation, which he defines as: "innovation that is only new to the countries and the firms that adopt the new product, process or organization, but is not necessarily new to the world, and is sometimes already known by consumers, in one form or another, in more backward countries" (Niosi 2012, p. 3). Niosi further argues that radical innovations causing structural changes on a global scale are rare; incremental innovations, on which imitation is based, are instead the norm.

Niosi (2012) formulates three general propositions. (1) Technological and organizational imitation is universal in economic development and has been theoretically underestimated. As Bolton (1993) has emphasized, Western pathos admires innovation and downplays imitation. The recent paper by Luo, Sun and Le Wung (2011) aimed at "emerging country copycats", illustrates the same trend. (2) Organizational imitation is an integral part of the diffusion process. (3) Imitation most often involves a certain degree of technological or organizational innovation, and there is a high degree of continuity linking imitation and incremental innovation. He concludes this part of his paper by stressing that: "imitative strategies, with a few exceptions (…) have been too often overlooked. They deserve a more thorough analysis, both at the micro and macro levels, in the debates about catching up, learning, and economic development" (Niosi 2012, p. 7).

Our aim in this paper is to add a fundamental analysis of the imitative activity of economic agents to the debate. We may conceive of this in more abstract terms by reference to innovators who cause structural and innovative change in the production sphere of economies by gaining a monopoly profit from innovative production and commercializing new goods, services or technologies. This initial state of affairs allows us to distinguish between firms that are laggards and firms that are leader-innovators. It should be borne in mind that the former are also heterogeneous agents, and that some of them are "preferred" as imitators in the copying of the production plans of leader-innovators. As a consequence, these imitators become secondary innovators operating in a production subsystem that is deprived of previous, primary innovators. This special selection mechanism for innovators that are once, twice or three times removed (extending potentially to infinity) from the primary innovators determines the diffusion of innovations, which is what guarantees their market success. Indeed, using a metric of innovation, we test the hypothesis that the more imitators there are, the more intensive are the innovative changes.

In the sense that they are reduced to, and operate in, a subsystem of a given economy, the paper seeks to study the logic of the imitative processes that define secondary innovators. This logic can be reconstructed by reference to the research programme on modelling Schumpeterian innovative evolution within the Arrow-Debreu dynamic general equilibrium theory (Malawski 1999, Malawski & Woerter 2006, Ciałowicz & Malawski 2011, *Innovative Economy* 2013), for which this framework would appear to provide an effective and convenient toolkit. Indeed, economic development in the Schumpeterian sense is modelled in this approach by innovative extension of the production system as a component of the Debreu economy, which is a setting that can serve as the base for studying imitative processes. The present study will therefore analyse the internal structure of the production system in a static setting and, specifically, explore the central hypothesis – that the more imitators there are, the more intensive are the innovative changes – axiomatically in the form of a theorem.

The paper examines a production system and an innovative extension of a production system in Section 2, before setting out an axiomatic analysis of a process of imitation in the given system in Section 3. The paper then proceeds to a study by theorems of the influence of imitation and imitators on the innovativeness of the production system. Conclusions are drawn in Section 5.

2. The Production System and Its Innovative Extension

The formal model of a production system takes the form of a two-range relational system (Debreu 1959, Malawski 1999, *Innovative Economy* 2013):

$$P = (B, \mathbb{R}^{\ell}; y, p, \eta, \pi),$$

where:

$B = \{b_1, \ldots, b_n\}$ is a finite set of the producers,

$\mathbb{R}^{\ell}$ is an ℓ-dimensional commodity-price space,

$y \subset B \times P(\mathbb{R}^{\ell})$ is a correspondence of production sets that to every producer $b \in B$ assigns a production set $y(b) = Y_b \subset \mathbb{R}^{\ell}$ that is a non--empty subset of the commodity space and represents the producer's feasible production technology,

$p \in \mathbb{R}^{\ell}$ is a price system,

$\eta \subset B \times P(\mathbb{R}^{\ell})$ is a correspondence of supply that to every producer $b \in B$ assigns a set $\eta(b)$ of the production plans maximizing his profit py_b in a price system p; that is to say: $\eta(b) := \eta_b(p) := \{y'_b \in Y_b\colon py'_b = \max_{y_b \in Y_b} py_b\}$,

$\pi : B \to \mathbb{R}$ is a maximum profit function that measures the maximum profit value in the set of plans $\eta(b)$, that is to say for $b \in B$: $\pi(b) := \pi_b(p) := \max_{y_b \in Y_b} py_b$.

In short, the production system is denoted: $P = (B, \mathbb{R}^{\ell}; Ch_P)$ where $Ch_P = (y, p, \eta, \pi)$ is a characteristic of system P.

In this system, each producer $b \in B$, operating in an ℓ-dimensional commodity-price space $\mathbb{R}^{\ell}$ tries to choose the production plans that will maximize profit in a given price system $p = (p_1, \ldots, p_{\ell}) \in \mathbb{R}^{\ell}$. The activities of a producer b which are governed by a set of production plans Y_b representing the producer's feasible production technology with respect to a correspondence of production sets y and by a feasible production plan, take the form $y_b = (y_{b1}, \ldots, y_{b\ell}) \in Y_b$. According to a correspondence of supply η and a maximum profit function π, which measures the maximum profit value in the set of plans $\eta(b)$ producers aim to select and execute the production plan that maximizes profits within the given price system.

Definition 2.1 (Ciałowicz & Malawski 2011, *Innovative Economy...* 2013)

Production system $P' = (B', \mathbb{R}^{\ell'}; y', p', \eta', \pi')$ can be called an innovative extension of system $P = (B, \mathbb{R}^{\ell}; y, p, \eta, \pi)$ (in short: $P \subset_i P'$), if

1) $\ell \leq \ell'$
2) $p = \text{proj}_{\mathbb{R}^{\ell}}(p')$
3) $\exists b' \in B' \quad \forall b \in B$
   (3.1) $\text{proj}_{\mathbb{R}^{\ell}}(Y'_{b'}) \not\subset Y_b$
   (3.2) $\text{proj}_{\mathbb{R}^{\ell}}(\eta'_{b'}(p')) \not\subset \eta_b(p)$
   (3.3) $\pi_b(p) < \pi'_{b'}(p')$.

According to this definition, production system P' is an innovative extension of system P if at least one new product or commodity can appear in P' (condition 1). These new products, which can be interpreted as a better way of meeting the needs present earlier in system P, are introduced by new firms or by firms that already exist. In production system P' there is at least one producer b' whose technological abilities exceed those of all the producers acting in production system P (condition 3.1). It follows that the optimal, profit-maximizing production plans of producer cannot be reduced to the analogous plans being executed by the producers b' in production system P (condition 3.2). What is more, although the prices of "old" products do not change (condition 2), the maximum profit a fixed producer can make is greater than that of any of the producers in system P (condition 3.3).

It is evident that when $\ell < \ell'$, Definition 2.1 covers at least four cases of the five internal changes that Schumpeter (1934, p. 66) defines as development:

1) the introduction of a new good – condition 1,
2) the introduction of a new method of production – condition 3.1,
3) the opening of a new market – condition 1,
4) the reorganization of an industry – condition 3 as a whole.

Remark 2.1

1. A producer $b' \in B'$ that satisfies conditions 3.1, 3.2 and 3.3 is called an innovator. The set of all innovators is denoted by B'_{in}.

2. Some innovative production plans that satisfy condition 3.2 can be found among the new production plans of an innovator b' defined by condition 3.1. Innovator b' maximizes its profit, which is greater than that of any of the producers in system P.

3. The innovative production plans of innovator b' in production system P' are compared to respective characteristics of system P. At the same time,

the structure of the set of innovators is neglected. Of course, if there is an innovator $b' \in B'$ under conditions 1 and 2 then $P \subset_i P'$.

4. Conditions 1 and 2 are formally independent. If they obtain simultaneously, new commodities cannot appear as manna from heaven, that is, in complete isolation from the previous technological structure, which is modified by innovative production plans.

Formally: if $y'_{b'} = (y'_1, y'_2, \ldots, y'_{\ell'}) \in Y'_{b'}$ is an innovative plan condition 3.1 implies that $\mathrm{proj}_{\mathbb{R}^\ell}(y'_{b'}) \notin Y_b$ so $\forall\, y_b = (y_1, y_2, \ldots, y_\ell) \in Y_b$ $\exists\, k \in \{1, 2, \ldots, \ell\}$: $y'_k \neq y_k$. Hence innovative changes occur in the production of at least one commodity k. Moreover, for $\bar{k} \neq k$ $y'_{\bar{k}} = y_{\bar{k}}$.

5. The strict version of condition 1, $\ell < \ell'$, means that the radical innovations occur in the form of at least one completely new good or service, whereas $\ell = \ell'$ corresponds to incremental technological innovations.

The latter case would appear to be common in practice.

Proposition 2.1

If $P \subset_i P'$, $\ell = \ell'$, $y'_{b'}$ is an innovative production plan and there exists a unique (in short: $\exists!$) $k \in \{1, 2, \ldots, \ell\}$ such that $y'_k \neq y_k$ and $p_k > 0$ (this commodity is a rare good), then $y'_k > y_k$.

Proof

According to Remark 2.1 (4), condition 3.1 of Definition 2.1 means that $\exists\, y'_{b'} = (y'_1, y'_2, \ldots, y'_{\ell'}) \in Y'_{b'}$ $\forall\, y_b = (y_1, y_2, \ldots, y_\ell) \in Y_b$ $\exists!\, k \in \{1, 2, \ldots, \ell\}$: $y'_k \neq y_k$.

Moreover, condition 3.2 implies $y'_{b'} \in \eta'_{b'}(p')$, so $\pi'_{b'}(p') = p' \cdot y'_{b'} = p'_1 y'_1 + p'_2 y'_2 + \ldots + p'_{\ell'} y'_{\ell'}$.

If from condition 3.3 we have $\pi_b(p) < \pi'_{b'}(p')$, so $\forall\, y_b = (y_1, y_2, \ldots, y_\ell) \in Y_b$:

$$p \cdot y_b \leq \pi_b(p) < p \cdot y'_{b'} \Leftrightarrow p_1 y_1 + p_2 y_2 + \ldots + p_\ell y_\ell < p_1 y'_1 + p_2 y'_2 + \ldots + p_\ell y'_\ell$$

and consequently $p_k y_k < p_k y'_k$, then for $p_k > 0$: $y'_k > y_k$.

Remark 2.2

1. There are two different effects of innovative changes in technologies. Let $k \in \{1, 2, \ldots, \ell\}$. Then:

a) If commodity k, such that $y'_k \neq y_k$ is an output (a positive coordinate in a production plan) in an innovative production plan and other coordinates are fixed, then condition $y'_k > y_k$ means that the level of production of this commodity has increased.

b) If commodity k is an input, the condition $y'_k > y_k$ means that the technology used to produce this commodity is more efficient.

2. It is possible that for all the commodities that are different from commodity k $(\hat{k} \neq k)$ condition $p \cdot y_b \leq \pi_b(p) < p \cdot y'_{b'}$ implies that $p_{\hat{k}} y'_{\hat{k}} < p_{\hat{k}} y_{\hat{k}}$. Yet with the standard assumption that $p_{\hat{k}} > 0$, we have $y'_{\hat{k}} < y_{\hat{k}}$. This means that:

a) If commodity $\hat{k}$ is an output, its level of production decreases. This is because it is less innovative than commodity k (if commodity k is an output) or it is displaced from the market by any other commodity (if commodity k is an input).

b) If commodity $\hat{k}$ is an input in the production of another product, this technology is less efficient than before.

It is possible to generalize Proposition 2.1 to a case in which there are more commodities $k \in \{1, 2, \ldots, \ell\}$ in the given innovative production plan for which $y'_k \neq y_k$.

Proposition 2.2

If $P \subset_i P'$, $\ell = \ell'$, $y'_{b'}$ is an innovative production plan and for $k \in \{1, 2, \ldots, \ell\}$ such that $y'_k \neq y_k$ there is $p_k > 0$, then $y'_k > y_k$.

This proof is similar to the proof of Proposition 2.1.

3. Imitation as Secondary Innovation

Schumpeter (1934) defined technological change as having three main interrelated stages: invention (producing new ideas), innovation (implementing new ideas in products and processes) and the diffusion of innovation based on imitations (the spread of new technology among its potential uses). It is consistent with empirical observation that a wave of imitative activity follows a creative innovation. Schumpeter wrote: "(...) if anyone has in him all that pertains to success (...) then he (the innovator) can make a profit which remains in his pocket. But he has also triumphed for others, blazed the trail and created a model for them which they can copy. They can and will follow him, first individuals and then whole crowds" (1934, p. 133). This means that imitation allows firms to adopt techniques from other firms that they are not yet using in their production processes. So it is that, in spite of the extraordinary outpouring of innovative products and new technologies that we are witnessing today, imitation generates a far greater flow of novelty than innovation.

We need only look around us to see that imitation is not only more abundant than innovation, but actually a much more prevalent pathway to business growth and profits. In the real world, companies copy and succeed. IBM, Texas Instruments and Holiday Inns got into computers, transistors and motels as imitators. Just as the iPod was not the first digital-music player, the iPhone was not the first smartphone and the iPad not the first tablet. Apple imitated products but made them far more appealing. The pharmaceutical industry is split between inventors and imitators, while the multi-billion-dollar market for supermarket own-label products is based on copying well-known brands. In some cases this even extends to copying the packaging. The pace and intensity of legal imitation has quickened in the last twenty years. Rather than implying clones of goods or illegal counterfeits, global competition shows us that legal imitation can be a very positive force in a firm's development.

We can study innovative changes by investigating imitation in the demand and supply sides of an economy and by distinguishing a set of imitators among producers and consumers. In the context of the Schumpeterian approach, we will concentrate not on copying itself, but rather on imitators of innovative production plans. In this way we will study how imitation affects the supply side and can drive the diffusion of imitation. We begin by defining an imitative extension of a production system.

Let three production systems be given:

$$P = (B, \mathbb{R}^{\ell}; y, p, \eta, \pi), P' = (B', \mathbb{R}^{\ell'}; y', p', \eta', \pi'), P'' = (B'', \mathbb{R}^{\ell''}; y'', p'', \eta'', \pi''),$$

where $p'' = p', \ell \leq \ell' = \ell'', B = B' = B'', P \subset_i P'$ such that $B'_{in} \subset B'$ is a set of producer-innovators.

Definition 3.1

A production system P'' is called an imitative extension of a production system P' in short: $P' \subset_{ii} P''$) if there exist producers $b'_i \in B'_{in}$, $b'' \in B''$, $b'' \neq b'_i$ with production plans $y'_{b'_i} \in Y'_{b'_i}$, $y''_{b''} \in Y''_{b''}$ such that $y''_{b''} = y'_{b'_i}$ and $y'_{b'_i}$ is an innovative plan.

This definition is consistent with Niosi's concept of imitative innovation (2012).

Remark 3.1

1. Production plan $y''_{b''}$ is called an innovative imitation of plan $y'_{b'_i}$.
2. Producer b'' is called an imitator of producer b'_i.
3. If $y''_{b''}$ is an imitation of plan $y'_{b'_i}$ then $y''_{b''} \in Y''_{b''} \bigcap Y'_{b'_i} \neq \varnothing$.

Let $B''_{im} := \{b'' \in B''$: $\exists\, b'_i \in B'_{in}$ b'' is an imitator of the producer b'_i, $b'' \notin B'_{in}\}$ be a set of producer-imitators.

It is now easy to establish conditions to guarantee that production system P'', an imitative extension of P', is an innovative extension of production system P, $P \subset_i P''$. The following theorem is true.

Theorem 3.1

Let $P \subset_i P'$ and $P' \subset_{ii} P''$.

If there are producers $b' \in B'_{in} \subset B'$, $b'' \in B''$, with production plans $y'_{b'} \in Y'_{b'}$, $y''_{b''} \in Y''_{b''}$, such that:

1) $y'_{b'}$ is an innovative production plan (Definition 2.1)

2) if production plan $y''_{b''} \in \eta''_{b''}(p'')$ is an imitation of innovative plan $y'_{b'}$ (Definition 3.1), then $y''_{b''}$ is an innovative plan with respect to production systems P and $P \subset_i P''$.

Proof

Let $y'_{b'} \in Y'_{b'}$ be an innovative production plan. This means that for each $b \in B$ $y'_{b'} \in (\text{proj}_{\mathbb{R}^\ell}(Y'_{b'}) \backslash Y_b)$, $y'_{b'} \in \eta'_{b'}(p')$ and $\pi_b(p) < \pi'_{b'}(p') = p'y'_{b'}$.

By Definition 3.1, if $y''_{b''}$ is an imitation of innovative plan $y'_{b'}$, then $y''_{b''} = y'_{b'}$. Moreover, for each $b \in B$ $y''_{b''} \in (\text{proj}_{\mathbb{R}^\ell}(Y''_{b''}) \backslash Y_b)$, $y''_{b''} \in \eta'_{b''}(p')$ and $\pi_b(p) < \pi'_{b''}(p') = p''y''_{b''} = p'y'_{b'}$. This means that $y''_{b''}$ is an innovative plan with respect to system P and, following Remark 2.1 (3), $P \subset_i P''$.

The next theorem demonstrates the relationship between imitator and innovator.

Theorem 3.2

Let $P \subset_i P'$, $P' \subset_{ii} P''$, $B = B' = B''$ and $b'_i \in B'_{in}$ be an innovator. If there exists $b'' \in B'' \backslash B'_{in}$ as an imitator of producer b'_i and $y''_{b''} \in \eta''_{b''}(p'')$ is an imitation of innovative plan $y'_{b'}$, then b'' is an innovator in production system P''.

Proof

Let $b'_i \in B'_i$ be an innovator, and $b'' \in B'' \backslash B'_i$ an imitator of producer b'_i. This means that there exists $y'_{b'} \in Y'_{b'}$ such that $y'_{b'}$ is an innovative plan, and there exists $y''_{b''} \in \eta''_{b''}(p'')$ such that $y''_{b''} = y'_{b'}$. Following theorem 3.1, $y''_{b''}$ is an innovative plan in system P'', so b'' is an innovator with respect to production system P.

Conclusion 3.1

$B''_{im} \subset B''_{in}$. This means that each imitator b'' in production system P'' for whom $y''_{b''} \in \eta''_{b''}(p'')$ is an imitation of innovative plan $y'_{b'}$ is an innovator in system P'' with respect to production system P. In other words, an imitator whose imitation production plan maximizes profits is a secondary innovator.

4. Imitation as a Driver of Innovativeness in Schumpeterian Perspective

Evaluating the innovativeness of an economy is one of the most important and difficult problems involved in analysing innovation processes. In this section we are concerned with comparing the innovativeness of two extensions of a production system. To do this we apply a metric of innovation (*Innovative Economy* 2013) that takes account of the qualitative changes in specific elements of the given model that are important for its innovativeness. This is a useful tool when studying the interaction of imitative and innovative activities in the process of innovative development. The aim of this section is to prove that imitations can intensify innovative changes in the production system and play a role as drivers of innovativeness.

Let a production system $P = (B, \mathbb{R}^{\ell}; y, p, \eta, \pi)$ be given.

For this system interpreted as a basic model, let the set of all possible innovative extensions be denoted by $\boldsymbol{P^i}$:

$$\boldsymbol{P^i} = \{P^i : P \subset_i P^i\}.$$

Let two innovative extensions of the basic model: $P^i_1, P^i_2 \in \boldsymbol{P^i}$ be given such that:

$$P^i_1 = (B^1, \mathbb{R}^{\ell_1}; y^1, p^1, \eta^1, \pi^1), P^i_2 = (B^2, \mathbb{R}^{\ell_2}; y^2, p^2, \eta^2, \pi^2).$$

Definition 4.1

A mapping $\rho_i : (\boldsymbol{P^i} \cup \{P\}) \times (\boldsymbol{P^i} \cup \{P\}) \to \mathbb{R}$ such that:

$$\rho_i(P^i_1, P^i_2) :=$$

$$\begin{cases} 0 & \text{if} \quad P_1^i = P_2^i \\ |\ell_1 - \ell_2| & \text{if} \quad \ell_1 \neq \ell_2 \\ |card(B_{in}^1) - card(B_{in}^2)| & \text{if} \quad \ell_1 = \ell_2 \text{ and } card(B_{in}^1) \neq card(B_{in}^2) \\ |\pi^1(p^1) - \pi^2(p^2)| & \text{if} \quad \ell_1 = \ell_2,\ card(B_{in}^1) = card(B_{in}^2) \text{ and } \pi^1(p^1) \neq \pi^2(p^2), \end{cases}$$

where $\pi^1(p^1) = \sum_{b \in B^1} \pi_b^1(p^1) \neq \pi^2(p^2) = \sum_{b \in B^2} \pi_b^2(p^2)$,

is called the innovative metric.

This definition covers a broad spectrum of specific subcases of innovative changes. Radical product innovation occurs in the case of $\ell_1 \neq \ell_2$, which means that in the two innovative extensions given there are (1) different processes of creation and (2) that a good or service that is a new or improved version of a previous good or service has been introduced. Product innovation is ruled out in the case of $\ell_1 = \ell_2$ and $card(B_{in}^1) \neq card(B_{in}^2)$, but the sets of innovators are changing. We may note that the populations of innovators are the same in the last case. But because changes in new technologies are hidden behind different maximum profits, this condition can be interpreted as a process innovation.

The defined metric allows us to measure the difference between selected elements of two production systems so that the innovativeness of two innovative extensions of a given system can be measured in terms of their distance from the basic model.

Definition 4.2

A production system P_2^i is called:

1) an extension of system P that is at least as innovative as system P_1^i, in short: $P_1^i \measuredangle_i P_2^i$, iff $\rho_i(P_1^i, P) \leq \rho_i(P_2^i, P)$,

2) a more innovative extension of system P than system P_1^i, in short: $P_1^i \angle_i P_2^i$, iff $\rho_i(P_1^i, P) < \rho_i(P_2^i, P)$.

The metric defined above can now be used to describe the role of imitators in an innovative development. Indeed, it can be proved that innovative changes intensify when the number of imitators grows.

Theorem 4.1

Let:

1) $P_1'' = (B^1, \mathbb{R}^{\ell_1}; y^1, p^1, \eta^1, \pi^1)$, $P_2'' = (B^2, \mathbb{R}^{\ell_2}; y^2, p^2, \eta^2, \pi^2)$ be two different imitative extensions of production system $P' = (B', \mathbb{R}^{\ell'}; y', p', \eta', \pi')$, which is an innovative extension of production system $P = (B, \mathbb{R}^{\ell}; y, p, \eta, \pi)$ (in short: $P \subset_i P'$, $P' \subset_{ii} P_1''$, $P' \subset_{ii} P_2''$),

2) $p' = p^1 = p^2$, $\ell = \ell' = \ell_1 = \ell_2$,
3) $P' \not\subset_i P''_1$ and $P' \not\subset_i P''_2$,
4) $card\,(B^1_{im}) < card\,(B^2_{im})$

then $P''_1 \angle_i P''_2$.

Proof

According to Theorem 3.2, we may notice that $P \subset_i P''_1$ and $P \subset_i P''_2$. There are of course no innovators in system *P*: $card(B_{in}) = 0$.

Moreover, taking into consideration Conclusion 3.1 and the assumptions $P' \not\subset_i P''_1$ and $P' \not\subset_i P''_2$ we have $B^1_{im} = B^1_{in}$ and $B^2_{im} = B^2_{in}$. Indeed, assumptions guarantee that $B^1_{im} \supset B^1_{in}$ and $B^2_{im} \supset B^2_{in}$ hold.

Thus, with the assumption that $\ell = \ell' = \ell_1 = \ell_2$, we have $\rho_i(P''_1, P) = = |card(B^1_{in}) - card(B_{in})| = card(B^1_{in})$ and $\rho_i(P''_2, P) = |card(B^2_{in}) - card(B_{in})| = = card(B^2_{in})$.

Hence, $card(B^1_{in}) = card(B^1_{im}) < card(B^2_{in}) = card(B^2_{im})$, so $P''_1 \angle_i P''_2$.

In general terms, the theorem states that innovative change becomes more intensive as the number of imitators increases.

5. Conclusions

This paper discusses the imitative activity of producers in a Schumpeterian process of structural change. Its chief aim is to show that imitators can be regarded as secondary innovators and that increases in the numbers of innovators intensify innovative changes throughout the production system. According to our approach, the innovations of leader-innovators are diffused when their production is imitated by other producers who can be regarded as secondary innovators.

In defining the intrinsic logic of the diffusion of innovations, the paper presents a new perspective on the important role played by imitators in innovative development. What is more, its results can be generalized to the whole Debreu economy where, for example, imitation can be analysed in a similar setting for a consumption system.

Bibliography

The Active Consumer: Novelty and Surprise in Consumer Choice (1998) ed. M. Bianchi. London–New York: Routledge.

Andersen, E. S. (2009) *Schumpeter's Evolutionary Economics*. London: Anthem Press.

Bessen, J. and Maskin, E. (2009) "Sequential Innovation, Patents, and Imitation". *Journal of Economics* 40 (4): 611–35, https://doi.org/10.1111/j.1756-2171.2009.00081.x.
Bolton, M. K. (1993) "Imitation versus Innovation. Lessons to be Learned from the Japanese". *Organizational Dynamics* 21 (3): 30–45, https://doi.org/10.1016/0090-2616(93)90069-d.
Ciałowicz, B. and Malawski, A. (2011) "The Role of Banks in the Schumpeterian Innovative Evolution – An Axiomatic Set-up" in A. Pyka, F. Derengowski, and M. da Graca (eds) *Catching Up, Spillovers and Innovation Networks in a Schumpeterian Perspective*. Heidelberg–Dordrecht–London–New York: Springer, pp. 31–58.
Ciałowicz, B. and Malawski, A. (2012) "The Role of Households in the Schumpeterian Innovative Evolution – An Axiomatic Set-up". Paper to the 14th International Schumpeter Society Conference, University of Queensland, Brisbane, Australia, 2–5 July.
Ciałowicz, B. and Malawski, A. (2013) "Demand Driven Schumpeterian Innovative Evolution" in A. Malawski (ed.) *Innovative Economy as the Object of Investigation in Theoretical Economics*. Kraków: Cracow University of Economics Press.
Debreu, G. (1959) *Theory of Value.* New York: Viley.
Glass, A. J. (2010) *Imitation as a Stepping Stone to Innovation.* College Station: Department of Economics, Texas A&M University.
Hanusch, H. and Pyka, A. (2006) "Principles of Neo-Schumpeterian Economics". *Cambridge Journal of Economics* 31 (2): 275–89, https://doi.org/10.1093/cje/bel018.
Herrmann-Pillath, C. (2013) *Foundations of Economic Evolution*. Cheltenham, UK–Northampton, MA: Edward Elgar.
Innovative Economy as the Object of Investigation in Theoretical Economics (2013) ed. A. Malawski. Kraków: Cracow University of Economics Press.
Luo, Y., Sun, J. and Le Wung, S. (2011) "Emerging Economy Copycats". *Academy of Management Perspectives* 25 (2): 37–56.
Malawski, A. (1999) *Metoda aksjomatyczna w ekonomii* [The axiomatic method in economics]. Wrocław: Ossolineum.
Malawski, A. and Woerter, M. (2006) "Diversity Structure of the Schumpeterian Evolution. An Axiomatic Approach". Arbeitspapiere / Working Papers of the Swiss Institute for Business Cycle Research No. 153, October. Zürich: KOF.
Mukoyama, T. (2003) "Innovation, Imitation, and Growth with Cumulative Technology". *Journal of Monetary Economics* 50 (2): 361–80, https://doi.org/10.1016/s0304-3932(03)00005-9.
Nelson, R. R. and Winter, S. (1982) *An Evolutionary Theory of Economic Change*. Cambridge, MA: Belknap Press of Harvard University Press.
Niosi, J. (2012) "Innovation and Development through Imitation (In Praise of Imitation)". Presented to the meeting of the International Schumpeter Society, Brisbane, Australia, 2–5 July.
Safarzyńska, K. and van den Bergh, J. C. J. M. (2010) "Evolutionary Models in Economics: A Survey of Methods and Building Blocks". *Journal of Evolutionary Economics* 20 (3): 329–73, https://doi.org/10.1007/s00191-009-0153-9.
Schumpeter, J. A. (1934) *The Theory of Economic Development*. Cambridge: Harvard University Press.
Segerstrom, P. S. (1990) *Innovation, Imitation and Economic Growth*. Econometrics and Economic Theory Paper No. 8818. East Lansing: Michigan State University.

Shenkar, O. (2010) “Defend Your Research: Imitation is More Valuable Than Innovation”. *Harvard Business Review*, April: 28–29.

Abstract

Logika procesów imitacyjnych: imitacje jako wtórne innowacje – schumpeterowska analiza aksjomatyczna

W pracy zaproponowana została aksjomatyczna analiza działalności imitacyjnej producentów w schumpeterowskim procesie zmian innowacyjnych. Zgodnie z przedstawionym podejściem zmiany strukturalne w systemie ekonomicznym zostają zapoczątkowane przez działalność liderów w sferze produkcji, będących radykalnymi innowatorami, którzy inicjują proces dyfuzji innowacji, a ich strategie oraz wprowadzone innowacje mogą być powielane przez producentów będących imitatorami. W rezultacie imitatorzy ci stają się innowatorami „drugiego rzędu”, dyskredytując jednocześnie innowatorów „pierwotnych”. Zgodnie z tym głównym celem przedstawionej pracy jest wykazanie, że zwiększenie liczby imitatorów prowadzi do zintensyfikowania zmian innowacyjnych w całym systemie produkcji. Ponadto przedstawione ujęcie jest zgodne z programem badawczym dotyczącym modelowania schumpeterowskiej ewolucji innowacyjnej w aparacie pojęciowym ujętej dynamicznie teorii równowagi ogólnej Arrowa-Debreu.

Słowa kluczowe: imitacje, innowacje, system produkcji, analiza aksjomatyczna, ujęcie schumpeterowskie.

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 15 • 2016
ISSN 1642-168X
AOC, 2016; 15: 57–74
DOI: 10.15678/AOC.2016.1504

Sabina Denkowska

ASSESSING THE ROBUSTNESS TO AN UNOBSERVED CONFOUNDER OF THE AVERAGE TREATMENT EFFECT ON THE TREATED ESTIMATED BY PROPENSITY SCORE MATCHING*

Abstract

One of the serious drawbacks of observational studies is the selection bias caused by the selection process to the treatment group. Propensity Score Matching (PSM), which allows for the reduction of the selection bias when estimating the average treatment effect on the treated (ATT), is a method recommended for the evaluation of projects and programmes co-financed by the European Union. PSM relies on a strong assumption known as the Conditional Independence Assumption (CIA) which implies that selection into the treatment group is based on observable variables, and all variables influencing both the selection process and outcome are observed by the researcher. If this does not hold, the estimated effect may be not so much the result of the treatment as of the lack of balance of an unobserved confounder, which affects both the selection process and the outcome. Rosenbaum's sensitivity analysis allows researchers to determine how strong the impact of such a potential unobserved confounder on selection into treatment and the outcome must be to undermine conclusions about ATT estimated by PSM. Rosenbaum's primal and simultaneous approaches are applied in the paper to assess robustness to an unobserved confounder of the net effect of internships for unemployed young people with a maximum age of thirty-five (estimated with PSM) organized by one of the biggest district employment offices in Małopolska.

Sabina Denkowska, Cracow University of Economics, Department of Statistics, Rakowicka 27, 31-510 Kraków, Poland, e-mail: denkowss@uek.krakow.pl

* The author acknowledge the support from research funds, granted to the Faculty of Management of the Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

Keywords: Propensity Score Matching, sensitivity analysis, Rosenbaum's sensitivity analysis, labour market policy.
JEL Classification: C14, E24.

1. Introduction

One of the serious drawbacks of observational studies is the selection bias caused by the selection process to the treatment group. Propensity Score Matching (PSM), which allows for the reduction of the selection bias when estimating the average treatment effect on the treated (ATT), is a method recommended (see EC 2014, pp. 6–7) for the evaluation of projects and programmes co-financed by the European Union. Propensity Score Matching refers to matching control units to treated units based on propensity scores, which are estimated based on observed characteristics. In common with other matching methods, PSM relies on a strong assumption known as the Conditional Independence Assumption (CIA) which "implies that selection is solely based on observable characteristics, and that all variables that influence treatment assignment and potential outcomes simultaneously are observed by the researcher" (Caliendo & Kopeinig 2008). Critics argue that the main disadvantage of studies based on matching units, including the PSM method, is connected with not taking into account an important confounder, which influences both the selection process and the outcome. This objection is relevant not only at the design and data-gathering stages. It may be the case that an unobserved confounder *U* is unmeasurable or difficult to measure. If a confounder has not been taken into consideration during the matching process, the treatment and control groups may not be comparable. In this way, the estimated effect may not have been caused by the treatment but by the lack of balance[1] introduced by an unobserved confounder *U*, which affects both outcome and selection[2]. It is for this reason that Rosenbaum (2005, 2010) recommends that observational studies based on matching are complemented by sensitivity analysis, which assesses the robustness of the estimated treatment effect in respect of an unobserved confounder.

The paper applies the primal and simultaneous Rosenbaum approaches to assess the robustness in respect of unobserved confounders of the net

[1] The balancing of variables means the similarity of distributions understood as the lack of systematic differences in their distributions.

[2] This problem is non-existent in experimental studies based on randomization, which balances all observed and unobserved variables. The differences between the values of the outcome variables in experimental and control groups are thus the result of the treatment performed on units in the experimental group.

effect of internships (estimated with PSM) organized by one of the biggest district employment offices in Małopolska for unemployed people with a maximum age of thirty-five.

2. Propensity Score Matching

2.1 Notation, Definitions, Assumptions

Let X denote a vector of observable characteristics and let D denote treatment (exposure) ($D \in \{0, 1\}$), where D = 1 means that a unit was exposed to treatment and D = 0 means that a unit was not exposed to treatment. For each i-th unit from an N-element population only one of two results for outcome variable Y is possible:

$$Y_i = \mathrm{D} \cdot Y_i^1 + (1 - \mathrm{D}) \cdot Y_i^0 = \begin{cases} Y_i^0, \text{gdy D} = 0 \\ Y_i^1, \text{gdy D} = 1 \end{cases}. \tag{1}$$

The usual aim of evaluation studies is to estimate an average treatment effect on the treated (ATT), which makes it possible to decide whether the treatment is effective for treated units:

$$\tau_{ATT} = \mathrm{E}[Y^1 - Y^0 \,|\, \mathrm{D} = 1]. \tag{2}$$

The average treatment effect on the treated can be expressed as the following difference:

$$\tau_{ATT} = (\mathrm{E}[Y^1 | \mathrm{D} = 1] - \mathrm{E}[Y^0 | \mathrm{D} = 0]) - (\mathrm{E}[Y^0 | \mathrm{D} = 1] - \mathrm{E}[Y^0 | \mathrm{D} = 0]) \tag{3}$$

in which the subtrahend is a selection bias resulting, among others things, from a lack of balance between the observed (and unobserved) variables in a treatment group and a control pool.

Matching in PSM is based on the propensity score p, which is defined as the conditional probability of being treated for a given vector x of observed characteristics X (Rosenbaum & Rubin 1983):

$$p(\mathrm{x}) := P(\mathrm{D} = 1 | X = \mathrm{x}) = \mathrm{E}(\mathrm{D} | X = \mathrm{x}). \tag{4}$$

The underlying assumption of PSM is the Conditional Independence Assumption[3] (CIA) that treatment assignment D is independent of potential outcomes (Y^1, Y^0) conditional on a given vector of covariates X (in the notation of Rosenbaum & Rubin 1983):

$$(Y^1, Y^0) \perp \mathrm{D} | X. \tag{5}$$

[3] Also known as "ignorability" (Rubin 1978), "no hidden bias" or "unconfoundedness" (Caliendo & Kopeinig 2008).

Rosenbaum and Rubin (1983) show that if potential outcomes are independent of treatment conditional on vector x of covariates X, they are also independent of treatment conditional on the propensity score $p(\mathrm{x})$. That CIA is untestable and, moreover, easy to undermine in observational studies, may mean that questions are raised about the results obtained using PSM.

The second assumption of PSM is the common support assumption, which is also known as the overlap assumption (Caliendo & Kopeinig 2008). It can be written as follows:

$$0 < P(\mathrm{D} = 1 | X = \mathrm{x}) < 1 \text{ for all x in support of } X. \quad (6)$$

This means that each unit with the same vector x of observed characteristics X has some chance of being treated and some chance of not being treated.

Unconfoundedness and the overlap assumption both constitute a property known as the "strong ignorability of assignment", which is necessary[4] to identify the treatment effect (Rosenbaum & Rubin 1983).

2.2 Algorithm for Propensity Score Matching

In practice, propensity scores are usually estimated as the fitted probabilities of treatment derived from the estimated logistic model, in which treatment status is regressed on observed baseline characteristics X. All of the variables simultaneously influencing the selection process and the outcome should be included[5] in the estimated logistic model (Stuart 2010). In the case of PSM, the model is only a means to achieve the goal of balancing the variables. For this reason, attention should be focused on the model's capacity to balance the variables rather than on estimating its parameters (Caliendo & Kopeinig 2008, Stuart 2010). Next, a method for matching[6] the control group to the treatment group (on the basis of estimated propensity scores) is selected[7]. Because the effects of treatment should not

[4] For ATT, however, Heckman et al. (1998) show that the unconfoundedness assumption can be weakened to conditional mean independence (see also Abadie & Imbens 2006). The overlap assumption can also be weakened because the right inequality in formula (6) is a sufficient condition for identifying ATT (Caliendo & Kopeinig 2008, Strawiński 2014).

[5] To satisfy the assumption of conditional independence (Rubin & Thomas 1996).

[6] For details of methods for matching the control group and the different ways they can be applied (with or without replacement, 1:*k* matching, caliper and radius) see Caliendo and Kopeinig (2008) and Stuart (2010).

[7] The Nearest Neighbour Method with 1:1 matching is the commonest method employed in evaluations of the Polish labour market (Wiśniewski & Maksim 2013, Konarski & Kotnarowski 2007, Trzciński 2009).

be assessed unless the matching is satisfactory, the latter is evaluated by checking and, where necessary, by determining a region of common support and checking the balance of variables included in the estimated logistic model. More information about determining the region of common support and about the indicators and tests used for assessing the balance of variables is available in Caliendo and Kopeinig (2008), Stuart (2010), Strawiński (2008, 2014) and Denkowska (2015). If the balance of variables is found to be unsatisfactory, researchers should consider other methods for matching or return to estimating the model of logistic regression and introduce two-way interactions and/or higher-order terms (Stuart 2010, Caliendo & Kopeinig 2008). Unfortunately, the tedious process of searching for the model and the best matching method that will balance all of the variables, higher-order terms and interactions from the estimated logistic model will not necessarily be successful. This may mean that the CIA has failed (Smith & Todd 2005). If this is the case, alternative approaches to evaluation should be considered (Caliendo & Kopeinig 2008).

The estimation of the treatment effect should not be conducted until all of the variables, higher-order terms and interactions of covariates used in the model have been satisfactorily balanced (Rubin 2001).

3. Sensitivity Analysis with Rosenbaum's Approaches

3.1. General Remarks

Sensitivity analysis has been proposed to indicate the magnitude of a hidden bias that should be present to alter the conclusions of the study. The robustness of average treatment effects estimated with matching methods can be assessed with Rosenbaum's approaches. Gastwirth, Krieger and Rosenbaum (1998) distinguish primal, dual and simultaneous approaches, which differ with regard to finding the thresholds of the association between the unobserved confounder and the exposure and/or the outcome that would render the test statistics of the study inference insignificant (Liu, Kuramoto & Stuart 2013). In the primal approach, the sensitivity parameter Γ relates an unobserved confounder U to treatment D, while assuming that confounder U is a perfect predictor of the outcome. In the dual approach, sensitivity parameter Δ relates an unobserved confounder U to outcome Y, while assuming that a confounder U is a perfect predictor of the treatment. Simultaneous sensitivity analysis uses both sensitivity parameters: Γ and Δ. The primal and simultaneous procedures are the most important from a practical point of view.

Rosenbaum's sensitivity analyses assume that matching is performed without replacement.

3.2. Rosenbaum's Primal Approach

Assuming that an unobserved confounder is a perfect predictor of the outcome, the question to be answered in the primal approach is how strong its association with treatment has to be to change the conclusions of the study.

Let us assume that there is an unobserved covariate U ($U \in <0; 1>$).

In matching methods we assume that a matched pair of units $\boldsymbol{k}$ and $\boldsymbol{l}$ with the same characteristics X ($\mathbf{x}_k = \mathbf{x}_l$) have the same probability of receiving treatment ($\pi_k = \pi_l$). But because of a potential unobserved confounder U, the odds that unit $\boldsymbol{k}$ receives treatment *de facto* are:

$$\frac{\pi_k}{1-\pi_k} = \exp(\kappa(\mathbf{x}_k) + \gamma u_k), \text{ where } 0 \le u_k \le 1. \tag{7}$$

So, for two units $\boldsymbol{k}$ and $\boldsymbol{l}$ with the same characteristics X ($\mathbf{x}_k = \mathbf{x}_l$) the odds ratio is:

$$\frac{\frac{\pi_k}{1-\pi_k}}{\frac{\pi_l}{1-\pi_l}} = \frac{\exp(\kappa(\mathbf{x}_k) + \gamma u_k)}{\exp(\kappa(\mathbf{x}_l) + \gamma u_l)} = \exp(\gamma(u_k - u_l)). \tag{8}$$

Rosenbaum (2002) shows the following bounds on the odds ratio:

$$\frac{1}{\exp(\gamma)} \le \frac{\frac{\pi_k}{1-\pi_k}}{\frac{\pi_l}{1-\pi_l}} \le \exp(\gamma). \tag{9}$$

Let $\Gamma := exp(\gamma)$. The units with the same values of observed covariates may nonetheless differ in terms of an unobserved confounder, so that one unit has odds of treatment that are up to $\Gamma \ge 1$ times greater than the odds for another unit" (Rosenbaum 2002, 2005).

Sensitivity analysis to an unobserved confounder proposed by Rosenbaum (2002) is based on several different randomisation tests (Liu, Kuramoto & Stuart 2013, Keele 2010). For a binary[8] outcome variable Y, the sensitivity analysis is based on McNemar's test, which is used to verify whether the confounder U has a significant impact on the result of the outcome variable Y. Information on paired units is presented in a contingency table (2×2).

[8] For other outcomes, the sensitivity test is based on the Wilcoxon sign rank test and the Hodges--Lehmann point estimate for the sign rank test (Rosenbaum 2005, Keele 2010).

The chances of being selected for a treated group are theoretically the same for units paired based on propensity score. When we employ Rosenbaum's sensitivity analysis we are seeking the odds ratio of treatment of the paired units (occurring due to unobserved confounder U) that would change the conclusions of the study in such a way as to render it insignificant.

Let T denote the number of all pairs in which the results of the outcome variable Y differ, and let a denote the number of pairs in which a treated unit has a positive result for the outcome variable and a not-treated unit has a negative result. The lower and upper bounds on the p-value are calculated by analogy with the binomial test p-value:

$$p_{lower} = \sum_{i=a}^{T} \binom{T}{i} (p^{-})^{i} (1-p^{-})^{T-i} \text{ and } p_{upper} = \sum_{i=a}^{T} \binom{T}{i} (p^{+})^{i} (1-p^{+})^{T-i}, \tag{10}$$

where the probabilities:

$$p^{-} = \frac{1}{1+\Gamma} \text{ and } p^{+} = \frac{\Gamma}{1+\Gamma} \tag{11}$$

are lower and upper bounds on the probability of being treated and are determined for different, hypothetical values of Γ. The lower bound p_{lower} is always lower than the observed p-value and, thereby, less important and rarely taken into account. Calculations are repeated with different values of Γ to find the value of parameter Γ in which p_{upper} becomes greater than 0.05.

3.3. Rosenbaum's Simultaneous Approach

By analogy with the primal analysis (Gastwirth, Krieger & Rosenbaum 1998), the upper bound on the p-value, p_{upper} in formula (10) is calculated in Rosenbaum's simultaneous approach with (Liu, Kuramoto & Stuart 2013):

$$p^{+} = p(treated) \cdot p(outcome) + (1 - p(treated)) \cdot (1 - p(outcome)), \tag{12}$$

where

$$p(treated) = \frac{\Gamma}{1+\Gamma}, \tag{13}$$

$$p(outcome) = \frac{\Delta}{1+\Delta} \tag{14}$$

are determined for different, hypothetical values of Γ and Δ.

A combination of values of Γ and Δ for which $p_{upper} \geq 0.05$, is the point at which the result is sensitive to an unobserved confounder U (Liu, Kuramoto & Stuart 2013).

4. Application of Rosenbaum's Sensitivity Analysis to the Study on the Net Effect of Internships

The net effect of the internships for unemployed people with a maximum age of thirty-five organized by the Tarnów District Employment Office was estimated. The purpose was to gauge the general effectiveness of internships organized by district employment offices in activating people who are young and unemployed[9]. The study was conducted using PSM. Rosenbaum's primal and simultaneous sensitivity analyses were applied to the estimated effect to check its robustness to a potential unobserved confounder influencing both the inclusion to the group of interns and finding a job.

In 2013, 1,409 unemployed people with a maximum age of thirty--five began internships. They were completed at least three months before 10 August 2014. The data were obtained from the Syriusz computer system, which is used to register unemployment.

The *X* variables employed in the study can be divided into four categories[10]:

I. Socio-demographic and health variables: *plec* – sex, *wiek* – age in years, *s_w* – single parenthood, *n_p* – disability, education (*w_brak* – lack, *w_sp* – elementary, *w_gim* – junior high school, *w_zaw* – vocational, *w_sr* – high school, *w_pm* – post-high school, *w_w* – university).

II. Employment, educational activity and activity on the labour market: job – classification[11] (*gr00* – lack, *grX* – where *X* denotes the classification number), *staz_pr* – number of years in employment, *dl_bzr* – long-term unemployment (Yes/No), *szk* – training during the two years preceding the internship (Yes/No), *l_prop* – number of job offers during the last six months, *w_a* – indicator of activity (community work, intervention jobs, training, internships, public work) in the two years preceding the internship: 0 – no active days, 1 – up to 100 active days, 2 – up to 200 active days. And so forth.

[9] A study of the net effect of the internships for all unemployed people – regardless of age – can be found in Denkowska (2015, 2016).

[10] The preliminary selection of variables was based on the experience gained from the Alternatywa II project, which the team evaluated using PSM. The project was part of the latest edition of Phare SSG RZL 2003 (Trzciński 2009). The source of data was SI PULS. After consulting the employees of the District Employment Office, however, it became clear that – due to the limitations of the Syriusz system – not all of the features could and should be used in the study. The paper describes the final set of variables used in the study.

[11] Classification in accordance with the ordinance of the Minister of Labour and Social Policy of 27 April 2010 on the Classification of Occupations and Specializations for Labour Market Needs.

III. Relative motivation to look for a job: *pr_zas* – eligibility for unemployment benefit.

IV. Skills and abilities: *pr_B* – driving licence, category B, *angBG* – at least a good knowledge of English, *angSL* – basic knowledge of English, *j_n* – knowledge of German.

The outcome variable *Y* was employment three months after finishing the internship. It was assumed[12] that a person not registered on the day the data were checked was employed.

The control pool consisted of 11,568 young people with a maximum age of thirty-five who had not been involved in an activation in 2013. To establish the values of variables *X* and outcome variable *Y*, for each person from the control pool the date of "starting internship" was randomly selected (measuring the values of the *X* variables). The average duration of the internship was added next. A check of whether or not the person was registered in the database was performed three months after the date the internship was "completed" (the value of variable *Y*).

The logistic regression model, in which the dependent variable was participation in the internship, was estimated first. There followed numerous attempts to obtain the best possible balance of variables, including modifying the regression model by introducing interactions and squares of variables, and checking various matching methods without replacement[13]. The distributions of propensity scores in the group of interns and the control group were analysed to check the region of common support, which influenced the decision to use a matching method with a caliper. The balance of variables, interactions and squares of variables were checked using standardized mean difference, and with *t*-tests for means in the interns group and in the control group, each time before and after matching. In the case of continuous and discreet variables, the similarity of distributions in the interns group and in the control group was verified using the bootstrap KS test[14].

The best balance of variables was obtained for the logistic model to which interactions and the $wiek^2$ variable were introduced (Table 1). The Nearest Neighbour Method used in the study (1:1, without replacement and with

[12] According to the methodology used by WUP (the regional employment office) in Kraków.

[13] Rosenbaum's sensitivity analysis can be applied only to matching methods without replacement.

[14] A bootstrap version of the Kolmogorov-Smirnov test can be used for both continuous and discreet random variables (Abadie 2002, Sekhon 2011).

a caliper[15] (*caliper* = 0.5)) led to the removal of two interns, for whom there were no good matches in the control group.

Table 1 presents the standardized mean differences[16] obtained from the formulas:

$$SDiff_{before} = \frac{\overline{X}_T - \overline{X}_{CP}}{S_T} \cdot 100\%,\ SDiff_{after} = \frac{\overline{X}_{TM} - \overline{X}_{CM}}{S_T} \cdot 100\%, \qquad (15)$$

where: $\overline{X}_T, \overline{X}_{CP}$ denote the means in the treatment (interns) group and in the control pool before matching, $\overline{X}_{TM}, \overline{X}_{CM}$ denote the means in the treatment (interns) group and in the control group after matching, while S_T stands for standard deviation in the treatment (interns) group before matching. The analysis of standardized differences is based on checking whether the values of standardized differences for all variables (per modulus) decrease after matching, and whether the values obtained after matching can be considered satisfactory. A standardized mean difference of at or below 3%, or at or below 5%, is considered sufficient in the majority of empirical studies (Caliendo & Kopeinig 2008).

Table 1 presents the standardized differences before and after matching, the *p*-values from the *t*-tests for the means of all variables, the interactions and $wiek^2$ variable, and the *p*-values from the KS bootstrap test[17] for all continuous and discreet variables and interactions.

All of the standardized differences decreased after matching, and none exceeded (per modulus) 4.3%. The *t*-tests did not reveal significant differences between the means. The Smirnov-Kolmogorov bootstrap test "confirmed" that the distributions for the continuous and discreet variables were similar.

After establishing that all of the variables, interactions and the $wiek^2$ variable were balanced, the net effect of the internships was estimated as the difference between the percentage of employed interns and the percentage of employed persons in the control group. The net effect of the internships for unemployed people with a maximum age of thirty-five was 10.945% with a standard error (see Imbens & Ambadie 2006) of 1.87% (p = 5.1905e – 09).

[15] Rubin and Thomas (1996) recommend keeping the limit at 0.25*s* or 0.5*s* where: $s = \sqrt{\frac{s_T^2 + s_{CP}^2}{2}}$, and s_T^2 and s_{CP}^2 denote variance in the treatment group and in the control pool respectively. It is worth noting that some participants from the treatment group may remain unmatched.

[16] The dichotomous variables were treated as continuous variables and standardized mean differences were obtained from the same formulas (Stuart 2010).

[17] A bootstrap version of the Kolmogorov-Smirnov test makes it possible to test the similarity of the distributions of both continuous and discreet random variables (Abadie 2002, Sekhon 2011).

Table 1. Standardized Mean Differences, *P*-values from *T*-tests for Means, and *P*-values from the Kolmogorov-Smirnov Bootstrap Test Before and After Matching

Variable	Before			After		
	$SDiff_{before}$	Test *t* *p*-value	KS bootstrap *p*-value	$SDiff_{after}$	Test *t* *p*-value	KS bootstrap *p*-value
Plec	–41.907	< 2.22e – 16	–	–2.1782	0.49353	–
wiek	–30.870	< 2.22e – 16	< 2.22e – 16	–2.2438	0.52446	0.556
s_w	–18.797	2.476e – 10	–	2.6609	0.45812	–
n_p	–6.4606	0.02512	–	0.46946	0.89976	–
w_sp	–38.026	< 2.22e – 16	–	–3.9018	0.30348	–
w_gim	–45.542	< 2.22e – 16	–	2.8634	0.41111	–
w_zaw	–70.761	< 2.22e – 16	–	–2.298	0.43519	–
w_sr	9.5860	0.0006772	–	2.7338	0.37518	–
w_pm	8.7961	0.001535	–	–0.87107	0.81645	–
w_w	46.639	< 2.22e – 16	–	–0.86959	0.75514	–
gr00	1.6127	0.56733	–	1.2729	0.69139	–
gr1	–0,5808	0.8388	–	0.0000	1.00000	–
gr2	45.166	< 2.22e – 16	–	–1.4605	0.60023	–
gr3	1.0624	0.7062	–	–0.21171	0.95084	–
gr5	–18,362	2.5091e – 10	–	2.7545	0.44239	–
gr6	–10.601	0.00058511	–	2.3879	0.47954	–
gr7	–59.703	2.22e – 16	–	1.5431	0.61887	–
gr8	–17.722	4.8343e – 09	–	–1.194	0.73892	–
gr9	–20.829	1.5147e – 11	–	–1.2311	0.74564	–
staz_pr	–47.266	< 2.22e – 16	< 2.22e – 16	–1.726	0.59530	0.01*
dl_bzr	–3.5778	0.20556	–	3.0175	0.38321	–
l_prop	9.3155	0.00086544	< 2.22e – 16	–1.9491	0.57823	0.796
pr_zas	–5.1164	0.074724	–	0.0000	1.00000	–
w_a	–3.9585	0.1599	0.056	0.0000	1.00000	0.876
szk	7.0995	0.0098951	–	–1.9673	0.58626	–
pr_B	31.148	< 2.22e – 16	–	–3.4191	0.26518	–
angBG	35.752	< 2.22e – 16	–	0.89826	0.76306	–
angSL	8.3716	0.0030145	–	–4.2772	0.19080	–
j_niem	15.494	3.6026e – 08	–	2.5881	0.42853	–
wiek^2	–33.603	< 2.22e – 16	< 2.22e – 16	–2.3053	0.51359	0.85997
*gr1*plec*	1.6906	0.53474		0.0000	1.00000	–
*gr00*plec*	–13.428	3.2142e – 06	–	3.1465	0.31151	–
*gr00*w_a*	2.9061	0.29779	0.308	2.5506	0.44051	0.628
*gr00*w_sp*	–21.093	3.54e – 11	–	–0.84576	0.81858	–
*wiek*d_bezr*	6.5890	0.020431	0.004	2.6035	0.45437	0.766
*plec*s_w*	–20.003	2.3416e – 09	–	–2.6688	0.52713	–
*dl_bzr*w_sp*	–20.269	0.29959	–	0.0000	1.00000	–
*n_p*w_zaw*	–6.1322	0.037246	–	1.8898	0.59303	–
*gr3*s_w*	–7.5494	0.012138	–	–1.194	0.73892	–
*gr3*w_w*	3.5677	0.19516		0.89119	0.79629	–
*gr00*wiek*	–2.2709	0.42444	< 2.22e – 16	1.2302	0.70606	0.95
*gr2*w_w*	45.344	< 2.22e – 16		–1.462	0.59719	
*gr3*l_prop*	5.9199	0.032278	0.03	2.5361	0.47542	0.7
*gr5*staz_pr*	–24.615	6.2172e – 15	< 2.22e – 16	–1.579	0.65040	0.25
*gr1*pr_zas*	2.0151	0.45614	–	0.0000	1.00000	–

Table 1 cnt'd

Variable	Before			After		
	$SDiff_{before}$	Test *t* *p*-value	KS bootstrap *p*-value	$SDiff_{after}$	Test *t* *p*-value	KS bootstrap *p*-value
*w_w*gr3*	3.5677	0.19516	–	0.89119	0.79629	–
*gr5*staz_pr*	–24.615	6.2172e – 15	< 2.22e – 16	–1.579	0.65040	0.25

Note: * variable *st_pr* is a continuous variable, so the similarity of distributions was also verified with the classic Kolmogorov-Smirnov test, which failed to reject the null hypothesis about the similarity of distributions after matching ($p = 0.11908$).

Source: author's own calculations in Matching package in R.

Despite the enormous effort invested in making exhaustive use of the information in the Syriusz system, doubt arose as to whether the causality observed between participation in internships and employment was not in fact caused by an unobserved confounder. We may be almost certain, for example, that certain personality features, such as entrepreneurship or communication skills, have a strong impact on employment. The decisive question here is how strong the impact of the unobserved factor on the selection process and employment should be to render the results statistically insignificant.

In order to conduct sensitivity analysis using Rosenbaum's primal approach, the information for 1407 pairs is presented in the contingency table[18] (Table 2). The number of pairs in which the results of outcome variable Y were different to each other was 712 ($T = 433 + 279$), and the number of pairs in which only interns were employed was 433 (a).

Table 2. Contingency Table for Paired Individuals

Group		Interns		Sum
		Employment	Lack	
Control group	Employment	431	279	710
	Lack	433	264	697
Sum		864	543	1407

Source: author's own calculations in Matching package in R.

During the next stage, for hypothetical values of Γ, probabilities p^- and p^+ were calculated, which were then used to obtain lower bounds and upper

[18] We may note, incidentally, that the odds ratio is 1.562, which means that for unemployed people with a maximum age of thirty-five, the odds for getting a job are 1.562 times greater for interns than for non-interns. In other words, the internship increases the odds for securing employment 1.562 times.

bounds for p-value according to formulas (10) and (11). The results of the calculations for selected values of Γ are presented in Table 3.

Table 3 informs us that the largest value (to two decimal places) of parameter Γ for which the probability p_{upper} was lower than 0.05, was 1.36. This means that the odds of one person in a pair becoming an intern can be 1.36 times greater than those of the other person in a pair because of different values for the confounder U, which has a powerful influence on employment, but there is still strong evidence that internships have an impact on employment ($p = 0.04578$). On the other hand, when $\Gamma = 1.37$, the relationship between internships and employment is no longer significant ($p = 0.05581$). $\Gamma = 1.36$ indicates[19] moderate robustness to the occurrence of an unobserved variable U.

Table 3. Bounds for Selected Values of Γ

Gamma	Probability	
	p_{lower}	p_{upper}
1.00	0.0000	0.00000
1.10	0.0000	0.00000
1.15	0.0000	0.00005
1.20	0.0000	0.00042
1.30	0.0000	0.01128
1.35	0.0000	0.03719
1.36	**0.0000**	**0.04578**
1.37	**0.0000**	**0.05581**
1.38	0.0000	0.06740
1.39	0.0000	0.08066
1.40	0.0000	0.09568
1.50	0.0000	0.34337

Source: author's own calculations in Matching package in R.

The results obtained (Table 3) were confirmed by an analysis conducted using the rbounds R package (Keele 2010). In this package Rosenbaum's primal approach is available for binary, ordinal and continuous variables for the matching variant 1:k (Keele 2014).

In Rosenbaum's simultaneous approach we look for the smallest values[20] of parameters Γ and Δ for which $p_{upper} \geq 0.05$ (calculated from formulas (10), (12)–(14)). We thus obtain points (Γ, Δ), at which the result is sensitive to

[19] Values of Γ (in the primal version of Rosenbaum's approach) in the social sciences are usually from 1 to 2 (Keele 2010).

[20] To one decimal place or two decimal places.

an unobserved confounder U (Liu, Kuramoto & Stuart 2013). The results of Rosenbaum's simultaneous analysis are presented in Table 4.

We are informed by the results in Table 4 that for $\Delta = \Gamma = 2.25$, p_{upper} is 0.035. This means that one person in a pair may be 2.25 times more likely to become an intern, and 2.25 times more likely to gain employment, than the other because they have different values of U. Yet there remains strong evidence that internships have an impact on employment ($p = 0.035$). Given $\Delta = \Gamma = 2.3$, on the other hand, the association between internships and employment would no longer be significant ($p = 0.0530$).

Table 4. Results (p_{upper}) of the Simultaneous Approach for Different Values of Γ and Δ

		Δ							
		1.0	1.5	2.0	2.25	2.3	2.5	3.0	+∞
Γ	1.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1.5	0.0000	0.0000	0.0000	0.0001	0.0001	0.0003	0.0010	0.3434
	2.0	0.0000	0.0000	0.0026	0.0096	0.0120	**0.0257**	**0.0957**	0.9995
	2.25	0.0000	0.0001	0.0096	**0.0350**	0.0432	0.0876	0.2701	1.0000
	2.3	0.0000	0.0001	0.0120	0.0432	**0.0530**	0.1060	0.3136	1.0000
	2.5	0.0000	0.0003	**0.0257**	0.0876	0.1060	0.1986	0.4945	1.0000
	3.0	0.0000	0.0010	**0.0957**	0.2701	0.3136	0.4945	0.8334	1.0000
	+∞	0.0000	0.3434	0.9995	1.0000	1.0000	1.0000	1.0000	1.0000

Source: author's own calculations in R.

Analysis of Table 4 also leads to the conclusion that, for example, one person in a pair may be twice as likely to become an intern and 2.5 times more likely to gain employment than the other because of different values of U. However, internships have a significant impact on employment ($p = 0.0257$). Were we to have $\Gamma = 2$ and $\Delta = 3$, though, the causality between internships and employment would no longer be significant ($p = 0.0957$). By analogy, furthermore, one person in a pair may be 2.5 times more likely to become an intern, and twice as likely to secure employment than the other because they have different values of U, but there is still strong evidence that internships have an impact on employment ($p = 0.0257$). Were we to have $\Gamma = 3$ and $\Delta = 2$, on the other hand, the causality between internships and employment would no longer be significant ($p = 0.0957$).

The primal approach provides a more sensitive indication than the simultaneous approach because it assumes a perfect relationship between

Table 5. Results (p_{upper}) of the Simultaneous Approach for Different Values of Γ and Δ

		Δ								
		1.0	1.3	1.36	1.37	1.4	1.5	2.0	2.5	+∞
Γ	1.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0000**
	1.3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0113**
	1.36	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0458**
	1.37	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0558**
	1.4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	**0.0957**
	1.5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	**0.3434**
	2.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0257	**0.9995**
	2.5	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0257	0.1986	**1.0000**
	+∞	**0.0000**	**0.0113**	**0.0458**	**0.0558**	**0.0957**	**0.3434**	**0.9995**	**1.0000**	**1.0000**

Source: author's own calculations in R.

the unobserved variable U and the outcome variable Y. The results presented in Table 5 demonstrate that the primal analysis of sensitivity is a particular case of the simultaneous approach. The probabilities (bold) for different values of Γ when $\Delta \rightarrow +\infty$ (or values of Δ when $\Gamma \rightarrow +\infty$) are the same as those in Table 3.

5. Conclusions

Because researchers conducting observational studies can never be sure that all confounders have been taken into account, sensitivity analysis is very important. Rosenbaum (2005, 2010) recommends a two-stage procedure for studies of this kind. What may be termed "classical" matching, which involves propensity scores estimated based on observed characteristics, should always be complemented with sensitivity analysis to asses robustness to a potential unobserved confounder. This practice will help increase confidence in the results obtained in observational studies. Higher values of Γ and Δ indicate robustness of the estimated effect to an unobserved confounder, while smaller values tell us that the result is sensitive to deviations from unconfoundedness, and remind us to proceed with caution in our interpretation.

The paper has set out the results of an empirical application of Rosenbaums's primal and simultaneous sensitivity analyses to the net effect of internships (estimated with PSM) for unemployed young people with a maximum age of thirty-five organized by one of the biggest district employment offices in Małopolska. It is unfortunate that PSM-based analyses of the labour market in Poland (see e.g. Wiśniewski & Maksim 2013, Konarski & Kotnarowski 2007, Trzciński 2009) have not been complemented by sensitivity analyses. Had they been performed, it would have been possible to relate the results of this study to other, similar studies. This is not, however, a signal to abandon analyses of the robustness of the estimated results. Quite the contrary. Robustness analysis should be incorporated as an important element of all observational studies. If decision-makers are armed with knowledge of the robustness of the estimated results, they are better equipped to draw conclusions from these studies.

Bibliography

Abadie, A. (2002) "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models". *Journal of the American Statistical Association* 97 (457): 284–92, https://doi.org/10.1198/016214502753479419.

Abadie, A. and Imbens, G. W. (2006) "Large Sample Properties of Matching Estimators for Average Treatment Effects". *Econometrica* 74 (1): 235–67, https://doi.org/10.1111/j.1468-0262.2006.00655.x.

Caliendo, M. and Kopeinig, S. (2008) "Some Practical Guidance for the Implementation of Propensity Score Matching". *Journal of Economic Surveys* 22 (1): 31–72, https://doi.org/10.1111/j.1467-6419.2007.00527.x.

Denkowska, S. (2015) "Wybrane metody oceny jakości dopasowania w Propensity Score Matching" [Selected quality assessment methods in propensity score matching] in K. Jajuga and M. Walesiak (eds) *Taksonomia 24. Klasyfikacja i analiza danych – teoria i zastosowania* [Taxonomy 24. Classification and analysis of data – theory and application]. Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu No. 384. Wrocław: Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, pp. 60–74.

Denkowska, S. (2016) "Zastosowanie analizy wrażliwości do oceny wpływu nieobserwowanej zmiennej w Propensity Score Matching" [The application of sensitivity analysis in assessing the impact of an unobserved confounder in propensity score matching] in K. Jajuga and M. Walesiak (eds) *Taksonomia 27. Klasyfikacja i analiza danych – teoria i zastosowania* [Taxonomy 27. Classification and analysis of data – theory and application]. Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu No. 427. Wrocław: Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, pp. 66–75.

EC (2014) *The Programming Period 2014–2020. Guidance Document on Monitoring and Evaluation. European Regional Development Fund and Cohesion Fund. Concepts and Recommendations*, March. Bruxelles: European Commission.

Gastwirth, J., Krieger, A. and Rosenbaum, P. (1998) "Dual and Simultaneous Sensitivity Analysis for Matched Pairs". *Biometrika* 85 (4): 907–20, https://doi.org/10.1093/biomet/85.4.907.

Heckman, J. J., Ichimura, H., Smith, J. and Todd, P. (1998) "Characterizing Selection Bias Using Experimental Data". *Econometrica* 66: 1017–98.

Keele, L. (2010) *An Overview of Rebounds: An R Package for Rosenbaum Bounds Sensitivity Analysis with Matched Data*, personal.psu.edu/ljk20/rbounds%20vignette.pdf. Accessed: 15 January 2016.

Keele, L. (2014) *Package 'Rbounds'*, https://cran.r-project.org/web/packages/rbounds/rbounds.pdf. Accessed: 15 January 2016.

Konarski, R. and Kotnarowski, M. (2007) "Zastosowanie metody propensity score matching w ewaluacji ex-post" [The use of propensity score matching in ex-post evaluations" in A. Huber (ed.) *Ewaluacja ex-post. Teoria i praktyka badawcza* [Ex-post evaluation. Theory and research practice]. Warszawa: PARP.

Liu, W., Kuramoto, S. K. and Stuart, E. A. (2013) "An Introduction to Sensitivity Analysis for Unobserved Confounding in Non-Experimental Prevention Research". *Prevention Science* 14 (6): 570–80, https://doi.org/10.1007/s11121-012-0339-5.

Rosenbaum, P. R. and Rubin, D. B. (1983) "The Central Role of Propensity Score in Observational Studies for Casual Effects". *Biometrika* 70 (1): 41–55, https://doi.org/10.1093/biomet/70.1.41.

Rosenbaum, P. R. (2002) *Observational Studies*. New York: Springer.

Rosenbaum, P. R. (2005) "Observational Study" in B. S. Everitt and D. C. Howell (eds) *Encyclopedia of Statistics in Behavioral Science*, vol. 3. New York: John Wiley & Sons.

Rosenbaum, P. R. (2010) *Design of Observational Studies*. New York: Springer.

Rubin, D. (1978) "Bayesian Inference for Causal Effects: The Role of Randomization". *The Annals of Statistics* 6 (1): 34–58, https://doi.org/10.1214/aos/1176344064.

Rubin, D. (2001) "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation". *Health Services & Outcomes Research Methodology* 2: 169–88.

Rubin, D. and Thomas, N. (1996) "Matching Using Estimated Propensity Scores: Relating Theory to Practice". *Biometrics* 52 (1): 249–64, https://doi.org/10.2307/2533160.

Sekhon, J. S. (2011) "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R". *Journal of Statistical Software* 42 (7): 1–52, https://doi.org/10.18637/jss.v042.i07.

Smith, J. and Todd, P. (2005) "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?". *Journal of Econometrics* 125 (1–2): 305–53, https://doi.org/10.1016/j.jeconom.2004.04.011.

Strawiński, P. (2008) "Quasi-eksperymentalne metody ewaluacji" [Quasi-experimental evaluation methods] in A. Haber (ed.) *Środowisko i warsztat ewaluacji* [The environment and technique of evaluation]. Warszawa: PARP, pp. 193–220.

Strawiński, P. (2014) *Propensity Score Matching. Własności małopróbkowe* [Propensity score matching. Small sample properties]. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.

Stuart, E. A. (2010) "Matching Methods for Causal Inference: A Review and a Look Forward". *Statistical Science* 25 (1): 1–21, https://doi.org/10.1214/09-sts313.

Trzciński, R. (2009) *Wykorzystanie techniki propensity score matching w badaniach ewaluacyjnych* [The use of propensity score matching in evaluation research]. Warszawa: PARP, http://www.parp.gov.pl/index/more/13335. Accessed: December 2014.

Wiśniewski, Z. and Maksim, M. (2013) “Polityka rynku pracy w Polsce – wyniki badań ewaluacyjnych prowadzonych za pomocą metody propensity score matching” [Labour market policy in Poland – the results of evaluation research carried out using propensity score matching] in *Rola funduszy unijnych w rozwoju społeczno-gospodarczym regionu* [The role of EU funds in the region’s socio-economic development], Zeszyty Naukowe Uniwersytetu Szczecińskiego No. 753. Szczecin: Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, pp. 93–110.

Abstract

Ocena odporności na występowanie nieobserwowanej zmiennej przeciętnego efektu oddziaływania na jednostki poddane oddziaływaniu oszacowanego za pomocą Propensity Score Matching

Jedną z poważnych wad badań obserwacyjnych jest obciążenie selekcyjne spowodowane selekcją jednostek do grupy poddawanej oddziaływaniu. Metoda Propensity Score Matching (PSM), która umożliwia redukcję obciążenia selekcyjnego podczas szacowania przeciętnego efektu odziaływania na jednostki poddane oddziaływaniu (ATT), jest metodą coraz częściej zalecaną przy ewaluacji projektów oraz programów współfinansowanych przez Unię Europejską. PSM opiera się na mocnym założeniu, zwanym założeniem warunkowej niezależności (CIA), które implikuje, że selekcja do grupy poddawanej oddziaływaniu musi być oparta wyłącznie na zmiennych obserwowanych i że wszystkie zmienne wpływające na poddanie oddziaływaniu oraz na potencjalne wyniki zmiennej wyjściowej są obserwowane przez badacza. Jeżeli założenie to nie jest spełnione, to oszacowany efekt może być nie tyle wynikiem oddziaływania, co skutkiem braku zbalansowania nieuwzględnionej (nieobserwowanej) w badaniu zmiennej, która wpływa zarówno na proces selekcji, jak i zmienną wyjściową. Analiza wrażliwości Rosenbauma umożliwia badaczom ocenę, jak silny musiałby być wpływ takiej potencjalnej nieobserwowanej zmiennej na proces selekcji oraz na zmienną wyjściową, aby podważyć wnioski na temat efektu ATT oszacowanego za pomocą PSM. Podejścia podstawowe oraz jednoczesne Rosenbauma są zastosowane w artykule do oceny odporności na występowanie nieobserwowanej zmiennej, efektu netto staży dla młodych bezrobotnych w wieku do 35 roku życia (oszacowanego za pomocą PSM), zorganizowanych przez jeden z największych powiatowych urzędów pracy w Małopolsce.

Słowa kluczowe: Propensity Score Matching, analiza wrażliwości, metody analizy wrażliwości Rosenbauma, polityka rynku pracy.

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 15 • 2016
ISSN 1642-168X
AOC, 2016; 15: 75–89
DOI: 10.15678/AOC.2016.1505

Przemysław Jaśko

STATISTICAL ARBITRAGE: A CRITICAL VIEW

Abstract

Statistical arbitrage dynamics is driven by a stationary, autoregressive process known as mispricing. This process approximates the value in time of a portfolio weighted equally to the elements of a cointegration vector of the log-prices processes of related instruments. Statistical arbitrage involves taking either long or short positions on a portfolio according to predictions of mispricing. This paper offers a theoretical analysis of cointegration testing under the conditional heteroscedasticity of the innovations process. Cointegration testing is used in the procedure of searching for the log-price processes of the related instruments that will form a statistical arbitrage portfolio. We also investigate dynamic characteristics of the mispricing process, which is a linear combination (cointegration vector elements are coefficients of it) of related log-prices processes for which the (T)VECM-MGARCH model class is assumed. Under this model assumptions making precise predictions on mispricing process based on past realizations are difficult. This paper can be treated as a starting point for an empirical analysis of statistical arbitrage portfolio construction. Reference is made to theory to describe the challenges which can be faced in constructing a statistical arbitrage portfolio based on cointegration, in modelling the dynamics of mispricing, and in prediction where the innovation process is conditionally heteroscedastic.

Keywords: statistical arbitrage, cointegration, conditional heteroscedasticity, VECM-MGARCH, Breitung cointegration test.
JEL Classification: C320, C580.

1. Introduction. A General Description of the Problem of Statistical Arbitrage

Statistical arbitrage[1] is a form of quantitative trading method which can be classified as a long-short, market neutral and relative pricing strategy. It is based on the assumption that the log-prices of related financial instruments,

Przemysław Jaśko, Cracow University of Economics, Faculty of Management, Computational Systems Department, Rakowicka 27, 31-510 Kraków, Poland, e-mail: jaskop@uek.krakow.pl

[1] As developed in Burgess (2000).

such as a subgroup of index constituent stocks or a term structure of interest rates, are driven by a reduced number of common stochastic trends, and that there is an equilibrium relation between the log-prices of these instruments. Deviations from the levels suggested by the equilibrium relation, caused by idiosyncratic shocks to the log-prices of a particular instrument or subgroup of instruments, are subjected to reversion by arbitrageurs and the related log-prices tend towards new levels at which the equilibrium relationship is satisfied.

Assuming that the equilibrium relation is given by the linear function $\boldsymbol{\beta}'\mathbf{x}_t = 0$ of related log-prices in vector $\mathbf{x}_t$, the process of deviations (also called the mispricing process) defined as $\{y_t = \boldsymbol{\beta}'\mathbf{x}_t\}$ should be a stationary, autoregressive process. In this case, the vector $\boldsymbol{\beta}$ elements are taken as portfolio weights and the value of y_t represents an approximate[2] value of such a portfolio over time. A portfolio with a structure of this kind is known as a statistical arbitrage portfolio or a Beta portfolio. In statistical arbitrage theory $\{y_t\}$, which approximates portfolio value, is a stationary, autoregressive process. When the y_t value deviates from 0, it is expected to move towards zero, which is informed by the level of expected value conditional on the process past. Anyone observing positive – or negative – deviations can then take a short – or long – position in a statistical arbitrage portfolio and make a profit by taking the opposite position when equilibrium is subsequently restored.

We demonstrate in this article that using information only on the expected value of y_t, conditional on the process past, is not sufficient to precisely forecast future movements of y_t values. According to the stylized facts about financial log-return processes (and therefore of log--prices as their cumulative sums), their innovation processes (stochastic input processes to dynamic models) are characterized by conditional heteroscedasticity, which is often of the MGARCH or MSV type, and sometimes also by unconditional heteroscedasticity. Because of this, the same idiosyncratic shocks (innovations) that cause deviations of y_t from the equilibrium level also inflate future conditional variances and covariances of innovations. This is in turn reflected in increased conditional variance of y_t, which is a linear combination of log-prices as shaped by innovations. This increased conditional variance makes it difficult to precisely forecast future movements of y_t values – despite the autoregressive property of y_t.

In statistical arbitrage problem, when we treat the log-prices of related instruments (for example the daily closing log-prices of stocks) as belonging

[2] The approximation is derived in Chan (2011).

to the class of integrated processes (most frequently as $I(1)$ vector processes), cointegration is applied to describe the equilibrium relations between log-prices and the VECM model (including its extensions) as a tool for modelling the dynamics of the log-prices vector process. This process is driven by the common stochastic trends, which makes it a $I(1)$ process, and the $I(0)$ temporary component shaped by an error correction mechanism and the short term dynamics of log-returns (the first differences of log-prices).

This paper offers a theoretical analysis of cointegration testing under the conditional heteroscedasticity of the innovations process. We also investigate dynamic characteristics of the mispricing process, which is a linear combination (coefficients of this combination are equal to cointegration vector $\boldsymbol{\beta}$ elements) of related log-prices processes for which the (T)VECM-MGARCH model class is assumed.

This paper can be treated as a starting point for an empirical analysis in statistical arbitrage portfolio construction. Reference is made to theory to describe the challenges which can be faced in constructing a statistical arbitrage portfolio based on cointegration, in modelling the dynamics of mispricing and in prediction, under a conditionally heteroscedastic innovation process.

We first present a formal definition of statistical arbitrage trading strategy and then consider the impact of innovations with conditional heteroscedasticity on cointegration based statistical arbitrage ability, and their influence on cointegration testing according to the frequentist approach.

2. Statistical Arbitrage

We define (after Jarrow et al. 2012) statistical arbitrage as a zero initial cost, self-financing trading strategy with a discounted cumulative trading profit value $V(n) = \sum_{i=1}^{n} \Delta V(i)$ (also called investor's wealth) for which:

1. $V(0) = 0$,
2. $\lim_{n\to\infty} E^P[V(n)] > 0$,
3. $\lim_{n\to\infty} P(V(n) < 0) = 0$,
4. $\lim_{n\to\infty} \frac{Var^P[V(n)]}{n} = 0$ if $P(V(n) < 0) > 0 \ \forall n < \infty$.

According to this definition, the expected value of discounted cumulative value in statistical arbitrage trading must, asymptotically, be positive. Statistical arbitrage strategy is different from traditional deterministic arbitrage in that it can exhibit negative discounted cumulative value – with positive probability in intermediate finite times – under conditions where the time-averaged variance of cumulative value for infinite time tends to zero and, asymptotically, the probability of a negative value for a trading strategy is zero.

The proponents of statistical arbitrage (Jarrow et al. 2012) assume that the dynamics of the incremental trading profits of statistical arbitrage (investor's wealth) can be described by the process:

$$\Delta V(i) = \mu i^{\theta} + \sigma i^{\lambda} Z_i,$$

where $\{Z_i\} \sim iiN(0,1)$ or $\{Z_i\} \sim MA(1)$.

Inference, if the constructed trading strategy can be considered statistical arbitrage, is based on testing a conjunction of hypotheses on the parameters of an incremental trading profits process: $H_1\colon \mu > 0, H_2\colon \lambda < 0$ and $H3\colon\ \theta > \max\left\{\lambda - \frac{1}{2}, -1\right\}$. An empirical series of investor's wealth deriving from statistical arbitrage trading is used in the testing.

3. Cointegration, the Heteroscedasticity of Model Innovations and Statistical Arbitrage

Before considering cointegrated processes it is necessary to define integrated n-dimensional (vector) processes.

We call the n-dimensional process $\{\mathbf{x}_t\}$ integrated of order 0 the process: $\{\mathbf{x}_t\} \sim I(0) \overset{df}{\Leftrightarrow} \mathbf{x}_t = \sum_{i=0}^{\infty} \boldsymbol{\gamma}_i L^i \boldsymbol{\epsilon}_t$, where L is a lag operator, $\{\boldsymbol{\epsilon}_t\} \sim WN(\mathbf{0}, \boldsymbol{\Sigma})$ (n-dimensional white noise process) and $\sum_{i=0}^{\infty} \boldsymbol{\gamma}_i \neq \mathbf{0}$.

We call the n-dimensional process $\{\mathbf{x}_t\}$ integrated of order d ($d \in \mathbf{Z}$) the processes:

$$\{\mathbf{x}_t\} \sim I(d) \overset{df}{\Leftrightarrow} \{\Delta^d \mathbf{z}_t\} \sim I(0) \text{ and } \{\Delta^{d-1} \mathbf{z}_t\} \nsim I(0).$$

Let us now assume n-dimensional process $\{\mathbf{x}_t\} \sim I(1)$ given by the VAR(k) model:

$$\mathbf{x}_t = \sum_{i=1}^{k} \boldsymbol{\Pi}_i \mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t,\ t = 1, \dots, t,$$

with $\{\boldsymbol{\epsilon}_t\} \sim iiN(\mathbf{0}, \boldsymbol{\Sigma})$, represented equivalently by:

$$\Delta\mathbf{x}_t = \boldsymbol{\Pi}\mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i \Delta\mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t,\ t = 1, \dots, T,$$

where $\mathbf{\Pi} = \sum_{i=1}^{k} \mathbf{\Pi}_i - \mathbf{I}_n$, $\mathbf{\Pi}_i = -\sum_{j=i+1}^{k} \mathbf{\Pi}_j$, with the characteristic polynomial matrix $\mathbf{A}(z) = (1-z)\mathbf{I}_n - \mathbf{\Pi} z - \sum_{i=1}^{k-1} \mathbf{\Gamma}_i (1-z)z^i$.

Additionally we assume $|\mathbf{A}(z)| = 0$ for z such that $|z| > 1$ or $z = 1$. The number of unit roots $z = 1$, is exactly $n - r$. For $z = 1$ we have $|\mathbf{A}(1)| = |-\mathbf{\Pi}| = 0$, implying that $\mathbf{\Pi}$ has reduced rank: $rk(\mathbf{\Pi}) = r < n$. We can thus make factorization $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ where $\dim(\boldsymbol{\alpha}) = \dim(\boldsymbol{\beta}) = n \times r$ and $rk(\boldsymbol{\alpha}) = rk(\boldsymbol{\beta}) = r$.

For the processes $\Delta\mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \mathbf{\Gamma}_i \Delta\mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t$ and $\boldsymbol{\beta}'\mathbf{x}_t$ (which is an r-dimensional process) to have initial conditions such that both will be $I(0)$ processes, it is necessary and sufficient that $|-\boldsymbol{\alpha}_\perp'\dot{\mathbf{A}}(1)\boldsymbol{\beta}_\perp| = |\boldsymbol{\alpha}_\perp'\mathbf{\Gamma}\boldsymbol{\beta}_\perp| \neq 0$, where $\dot{\mathbf{A}}(1) = \frac{\mathrm{d}}{\mathrm{d}z}\mathbf{A}(z)\big|_{z=1}$, $\mathbf{\Gamma} = \mathbf{I}_n - \sum_{i=1}^{k-1} \mathbf{\Gamma}_i$, and $\boldsymbol{\alpha}_\perp$, $\boldsymbol{\beta}_\perp$ are respectively $n \times (n-r)$ matrices of orthogonal complements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, with rank $rk(\boldsymbol{\alpha}_\perp) = rk(\boldsymbol{\beta}_\perp) = n - r$.

When these conditions are met, the Johansen version of the Granger Representation Theorem (Johansen 1995) states that $I(1)$ process $\{\mathbf{x}_t\}$ is cointegrated of order $(1,1)$: $\{\mathbf{x}_t\} \sim CI(1,1)$ and can be equivalently represented as (for $t = 1, \ldots, T$):

$$\Delta\mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \mathbf{\Gamma}_i \Delta\mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t,$$

$$\mathbf{x}_t = \mathbf{C}\sum_{i=1}^{t} \boldsymbol{\epsilon}_i + \mathbf{C}_1(L)\boldsymbol{\epsilon}_t + \mathbf{A},$$

where $\mathbf{C} = \boldsymbol{\beta}_\perp (\boldsymbol{\alpha}_\perp'\mathbf{\Gamma}\boldsymbol{\beta}_\perp)^{-1}\boldsymbol{\alpha}_\perp'$, $\mathbf{C}_1(L)\boldsymbol{\epsilon}_t \sim I(0)$ and $\boldsymbol{\beta}'\mathbf{A} = \mathbf{0}$ ($\mathbf{A}$ is associated with the initial value).

The column vectors from the $\boldsymbol{\beta}$ matrix form the basis of a cointegration space which is the r-dimensional subspace of $\mathbf{R}^n$, where $0 < r < n$ and, for any vector $\mathbf{b} \in \mathrm{sp}(\boldsymbol{\beta})$, we have $\{\mathbf{b}'\mathbf{x}_t\} \sim I(0)$, because $\mathbf{b}'\mathbf{C} = 0$, specifically $\boldsymbol{\beta}'\mathbf{x}_t$ forms an r-dimensional $I(0)$ process.

Summarizing for $\{\mathbf{x}_t\} \sim CI(1,1)$, we have: $\{\mathbf{x}_t\} \sim I(1)$, $\{\Delta\mathbf{x}_t\} \sim I(0)$, $\{\mathbf{y}_t = \boldsymbol{\beta}'\mathbf{x}_t\} \sim I(0)$, additionally $\{\boldsymbol{\beta}_\perp'\Delta\mathbf{x}_t\} \sim I(0)$.

Once the related log-prices have been identified, the central problem in statistical arbitrage is to model and forecast the deviations process. When we assume $r = 1$ (a higher cointegration rank may suggest that the chosen group of assets includes some mutually exclusive subgroups of related log-prices), the deviations process is represented by a scalar process $\{y_t = \boldsymbol{\beta}'\mathbf{x}_t\}$, which is a stationary, autoregressive process.

Unfortunately, when heteroscedastic variance is present in y_t, the autoregressive property is not a sufficient condition for a precise directional forecast of y_t and hence for taking profitable positions on a Beta portfolio based on it.

To demonstrate this, let us make further assumptions that incorporate stylized facts about financial log-returns by extending the VECM model with the *iiN* innovations process.

For most financial log-returns, innovations processes $\{\boldsymbol{\epsilon}_t\}$ show conditional heteroscedasticity, which is usually modelled by one of the many MGARCH variants, and are no longer strict white noise processes. Innovations processes are composed of variables that are not correlated in time, but are not independent in time. Unconditional heteroscedasticity, caused for example by structural breaks that permanently increase the mean dispersion level from a particular moment in time, is also sometimes observed. The heteroscedastic innovations referred to above are embraced by a group of martingale difference sequence (MDS) processes.

Let us consider a VECM-MGARCH[3] model for the log-returns of related stocks with a $CI(1,1)$ cointegrated n-dimensional log-prices process, where $r = 1$ implies $\boldsymbol{\beta}$ composed of only one cointegrating vector. For ease of interpretation we assume that there are no short-term dynamics in the model i.e. $\boldsymbol{\Gamma}_i = \mathbf{0}$, $i = 1, \ldots, k-1$.

VECM-MGARCH model ($t = 1, \ldots, T$):

$$\left.\Delta\mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t \right\} \text{ VECM part}$$

$$\left.\begin{aligned} &\boldsymbol{\epsilon}_t = \mathbf{H}_t^{1/2}\boldsymbol{\eta}_t \\ &\mathbf{H}_t = \mathbf{H}(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1}, \ldots, \mathbf{H}_{t-1}, \ldots) \\ &\{\boldsymbol{\eta}_t\} \sim iid(\mathbf{0}, \mathbf{I}_n) \end{aligned}\right\} \text{ MGARCH part,}$$

where $\mathbf{H}_t = \mathbf{H}_t^{1/2}(\mathbf{H}_t^{1/2})'$ is the "square root" decomposition of $\mathbf{H}_t = [h_{ij,t}]_{i,j=1,\ldots,n}$, representing a covariance matrix in moment t conditional on the past of the process, $\mathbf{H}$ is a matrix function representing MGARCH, with some previous values of $\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}'_{t-j}$ and $\mathbf{H}_{t-j}$ as arguments, $\{\boldsymbol{\eta}_t\}$ is an n-dimensional process of independent standardized variables, having for example a multivariate normal distribution or a multivariate t-Student

[3] So that more general statements can be made, the variant of the MGARCH model is not precisely specified.

distribution with vector mean **0** and covariance $\mathbf{I}_n$ (n-dimensional unit matrix), but also with asymmetric counterparts of these distributions.

The deviation process (mispricing process) for this model, with cointegration rank $r = 1$ and cointegration vector $\boldsymbol{\beta} = [\beta_1 \dots \beta_n]'$, is a scalar process $\{y_t\}$ given by:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t = (1 + \boldsymbol{\beta}'\boldsymbol{\alpha})\boldsymbol{\beta}'x_{t-1} + \boldsymbol{\beta}'\boldsymbol{\epsilon}_t$$
$$y_t = \phi y_{t-1} + \varepsilon_t^y,$$

where $\phi = (1 + \boldsymbol{\beta}'\boldsymbol{\alpha})$, $\phi \in (-1,1)$ for $\{\mathbf{x}_t\} \sim CI(1,1)$ and $\varepsilon_t^y = \boldsymbol{\beta}'\boldsymbol{\epsilon}_t$.

The deviations process $\{y_t\}$ is in fact stationary and autoregressive, but let us investigate its properties, such as its expected value and variance conditional on the past of the process.

Let $\Psi_t = \sigma(\mathbf{x}_s, s \le t)$ be a σ-algebra generated by the process $\{\mathbf{x}_s\}$ up to moment t.

$$E(y_t \mid \Psi_{t-1}) = \phi y_{t-1}$$

$$V(y_t \mid \Psi_{t-1}) = V(\varepsilon_t^y \mid \Psi_{t-1}) = V(\boldsymbol{\beta}'\boldsymbol{\epsilon}_t \mid \Psi_{t-1}) =$$
$$= \sum_{i=1}^{n} \beta_i^2 V(\varepsilon_{ti} \mid \Psi_{t-1}) + 2\sum_{i=1}^{n}\sum_{j>i} \beta_i\beta_j Cov(\varepsilon_{ti}, \varepsilon_{tj} \mid \Psi_{t-1}) \Leftrightarrow$$
$$\Leftrightarrow V(y_t \mid \Psi_{t-1}) = \sum_{i=1}^{n} \beta_i^2 h_{ii,t} + 2\sum_{i=1}^{n}\sum_{j>i} \beta_i\beta_j h_{ij,t}.$$

The conditional variance form for y_t shows that in general conditions $\{\varepsilon_t^y\}$ is not given by a univariate GARCH model. The first component in y_t conditional variance $\sum_{i=1}^{n} \beta_i^2 h_{ii,t}$ is always positive and cumulates (with positive multipliers β_i^2) the conditional variances $h_{ii,t}$ of the univariate constituents of $\boldsymbol{\epsilon}_t$ from the innovations process, thus increasing the value of $V(y_t \mid \Psi_{t-1})$. The second component, which is twice $\sum_{i=1}^{n}\sum_{j>i} \beta_i\beta_j h_{ij,t}$, can – but does not have to – take negative values and, in some conditions, can reduce the level of conditional variance of y_t. The sign of the second component depends on the signs of parameters β_i, β_j and on the conditional covariances $h_{ij,t}$ for the constituents of $\boldsymbol{\epsilon}_t$.

These findings confirm that, because of increased conditional variance $V(y_{t+1} \mid \Psi_t)$, information about $E(y_{t+1} \mid \Psi_t)$ is not on its own a precise indicator of future y_{t+1} value movements. Moreover, the conditional distribution $\boldsymbol{\epsilon}_{t+1} \mid \Psi_t$ type and parameters strongly affect the conditional distribution of $y_{t+1} \mid \Psi_t$ as a linear combination of $\mathbf{x}_{t+1} \mid \Psi_t$ constituents.

If useful predictions are to be made in a case such as this, all of the information available on the conditional distribution of $y_{t+1} \mid \Psi_t$ should be exploited rather than selected parameters only. From the conditional distribution of $y_{t+1} \mid \Psi_t$ we can derive quantile forecasts or assess the probability of up or down movement from the current y_t value. Because of the complex shape of the conditional distribution $y_{t+1} \mid \Psi_t$, which can be asymmetric, and owing to the complicated relations describing its parameters, there may occur a situation in which $\text{sgn}[E(y_{t+1} \mid \Psi_t) - y_t]$ gives a specific direction for future movement while the information on the conditional distribution of $y_{t+1} \mid \Psi_t$ suggests that movement in the opposite direction is more probable. Here, the autoregressive tendency to revert, which was expected, is dominated by overdispersion and statistical arbitrage cannot be realized.

We have simulated a sample series of a length of $T = 1000$ simulated from VECM-DCC-GARCH ($n = 2$, $r = 1$ with a 2×1 cointegration vector $\boldsymbol{\beta}$; the model has no short-term dynamics) for $\mathbf{x}_t$, $\Delta\mathbf{x}_t$, $y_t = \boldsymbol{\beta}'\mathbf{x}_t$. Figure 1 shows scatter plot for $\mathbf{x}_t = (x_{t1}, x_{t2})'$, Figures 2 to 4 are plots of the time series concerned.

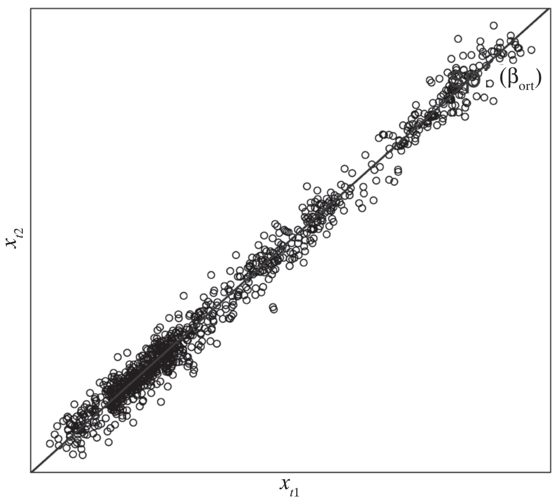

Fig. 1. Scatter Plot for $(x_{t1}, x_{t2})'$ with Marked Attractor Given by the Subspace $\text{sp}(\boldsymbol{\beta}_\perp)$

Source: author's own research.

The one-dimensional subspace spanned by the $\boldsymbol{\beta}$ orthogonal complement, denoted by $\text{sp}(\boldsymbol{\beta}_\perp)$, forms an attractor for process $\{\mathbf{x}_t\}$; as for $\mathbf{x}_t^* = c \cdot \boldsymbol{\beta}_\perp$ with arbitrary $c \neq 0$, we have $y_t^* = \boldsymbol{\beta}'\mathbf{x}_t^* = c \cdot \boldsymbol{\beta}'\boldsymbol{\beta}_\perp = 0$ and, for the assumed model, $y_t = \boldsymbol{\beta}'\mathbf{x}_t$ is driven towards 0.

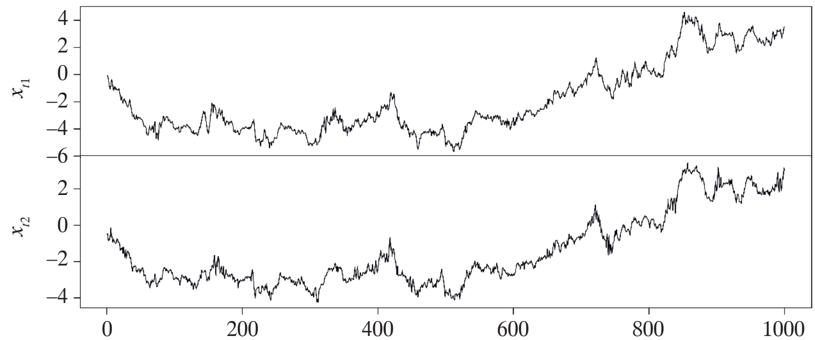

Fig. 2. Simulated Log-prices $\mathbf{x}_t = (x_{t1}, x_{t2})'$ Time Series

Source: author's own research.

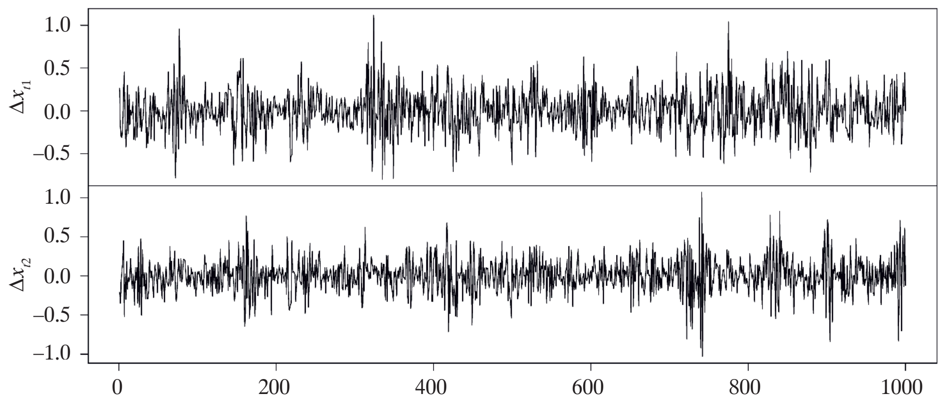

Fig. 3. Simulated Log-returns $\Delta\mathbf{x}_t = (\Delta x_{t1}, \Delta x_{t2})'$ Time Series

Source: author's own research.

The VECM-MGARCH may be too restrictive in its construction, since it is suggested that because of transaction costs, only higher absolute deviations from the equilibrium relation are corrected by arbitrageurs. An extension to the VECM part of the model, known as TVECM or Threshold VECM, was proposed to take account of this (Balke & Fomby 1997). In this case TVECM assumes three regimes and one cointegrating vector, $r = 1$:

$$\Delta\mathbf{x}_t = \sum_{m=1}^{3}\left(\boldsymbol{\alpha}^{(m)}\boldsymbol{\beta}^{(m)}{}'\mathbf{x}_{t-1} + \sum_{i=1}^{k-1}\boldsymbol{\Gamma}_i^{(m)}\Delta\mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t^{(m)}\right)\cdot I(c_{m-1} < y_{t-1} \leq c_m),$$

where I is an indicator function and for the middle regime $m = 2$ we have: $0 \in (c_1, c_2]$, $\boldsymbol{\alpha}^{(2)} \equiv 0$ which means there is no cointegration in the middle regime and $y_t = \boldsymbol{\beta}'\mathbf{x}_t \sim I(1)$ for $c_1 < y_{t-1} \le c_2$.

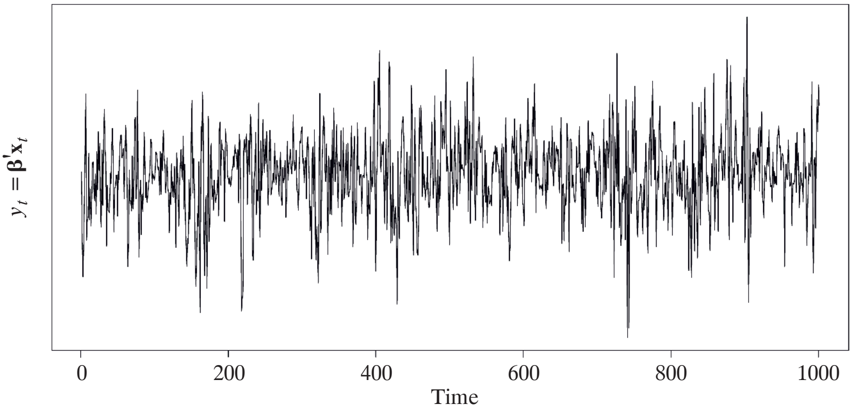

Fig. 4. Simulated Realization of Deviations (Mispricing) Process $y_t = \boldsymbol{\beta}'\mathbf{x}_t$

Source: author's own research.

In this case, because of nonlinear dynamics, the model does not have the representation stated by the Granger Representation Theorem.

To analyze the properties of vector processes with non-linear dynamics, concerning order of integration and cointegration, the definitions of integrated and cointegrated processes need to be extended.

The extended definition of the $I(0)$ n-dimensional (vector) process makes use of the functional central limit theorem (FCLT), whose formal aspects are described by Davidson (1994).

We call the n-dimensional process $\{\mathbf{x}_t\}$ an $I(0)$ process $\Leftrightarrow \forall\, a \in [0,1]$ and $T \to \infty$: $T^{-\frac{1}{2}} \sum_{t=1}^{\lfloor aT \rfloor} \mathbf{x}_t \xrightarrow{d} \boldsymbol{\Sigma}_{\mathbf{x}}^{1/2}\mathbf{W}(a)$, where d symbolizes weak convergence (convergence in distribution), $\lfloor \cdot \rfloor$ is a floor function, $\mathbf{W}(a)$ is an n-dimensional standard Wiener process, $\boldsymbol{\Sigma}_{\mathbf{x}} = \lim_{T \to \infty} T^{-1} Cov\left(\sum_{t=1}^{T} \mathbf{x}_t\right)$ is called a long-term covariance matrix and $\boldsymbol{\Sigma}_{\mathbf{x}}^{1/2}$ is its "square root" matrix. The definition of the processes for vector $I(d)$ remains unchanged.

In this extended approach, cointegration is defined without appealing to an explicitly specified model. In this way it can embrace models with different types of short-term and error-correction dynamics.

Let $\{\mathbf{x}_t\} \sim I(1)$ with respect to the extended definition. We additionally assume decomposition of the invertible matrix $\tilde{\boldsymbol{\beta}} = [\boldsymbol{\beta}_\perp, \boldsymbol{\beta}]$, where $\dim(\boldsymbol{\beta}) = n \times r$, $\dim(\boldsymbol{\beta}_\perp) = n \times (n-r)$, and $0 < r < n$ and $\boldsymbol{\beta}'\boldsymbol{\beta}_\perp = \mathbf{0}$.

Process $\{\mathbf{x}_t\}$ is $CI(1,1)$ if we can decompose it into two components: $\tilde{\boldsymbol{\beta}}'\mathbf{x}_t =$ $= \begin{bmatrix} \boldsymbol{\beta}_\perp' \\ \boldsymbol{\beta}' \end{bmatrix} \mathbf{x}_t = \begin{bmatrix} \mathbf{u}_t \\ \mathbf{y}_t \end{bmatrix}$, for which $T^{-\frac{1}{2}} \mathbf{u}_{\lfloor aT \rfloor} \xrightarrow{d} \mathbf{W}(a) \sim I(1)$ and $T^{-2} \sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}_t' = o_p(1)$, where $\mathbf{W}(a)$ is a $(n-r)$-dimensional standard Wiener process.

Here $\{\mathbf{y}_t = \boldsymbol{\beta}'\mathbf{x}_t\}$ represents a transitory component, which can also be generated by a nonlinear process with a short memory. In addition, $\boldsymbol{\beta}$ spans an r-dimensional cointegration space. $\{\mathbf{u}_t = \boldsymbol{\beta}_\perp'\mathbf{x}_t\}$, on the other hand, is a stochastic trend component, which is "variance dominating". This means that $\{\mathbf{u}_t\}$ diverges at a faster rate than $\{\mathbf{y}_t\}$.

4. Difficulties with Inference on Cointegration in the Case of Heteroscedastic Innovations

This paper offers a brief discussion of only the frequentist approach to testing cointegration under the heteroscedastic innovations of a specific type.

Classical Johansen cointegration rank tests associated with the $CI(1,1)$ process VECM model with *iiN* innovations, known as the maximum eigenvalue test (cointegration rank: H_0: r vs. H_1: $r+1$) and the trace test (cointegration rank: H_0: r vs. H_1: $n \Leftrightarrow \{\mathbf{x}_t\} \sim I(0)$), under the null hypotheses have asymptotic distributions, which are derived with the use of FCLT and specified as the functionals of the standard Wiener process.

It has been shown (Cavaliere, Rahbek & Taylor 2010) that when we attenuate assumptions about an innovations process from *iiN* to one that belongs to the MDS class of processes, which includes conditionally and unconditionally heteroscedastic processes, Johansen tests will weakly converge to the same asymptotic distributions.

In VECM models with heteroscedastic innovations, Johansen cointegration rank tests for finite-length samples are regarded as quasi-likelihood ratio tests because they use a likelihood function for the VECM model with *iiN* innovations. These Quasi-LR tests use asymptotic critical values, which is reflected in moderate to high test-size distortions. In a simulation study of Johansen tests using innovations with an MGARCH type of conditional heteroscedasticity (Maki 2013), a true null hypothesis of no cointegration ($r = 0$) was more frequently rejected than the nominal critical level assumed.

To improve the performance of the Johansen Quasi-LR tests for finite--length samples, a wild bootstrap procedure was suggested (Cavaliere, Rahbek & Taylor 2010). Unlike other bootstrap methods, such as the *iid* bootstrap (Swensen 2006), wild bootstrap makes it possible to retain the heteroscedasticity structure of the original series. In a single wild bootstrap replication, Quasi-Maximum Likelihood (QML) estimated VECM model errors $\{\boldsymbol{\epsilon}_t\}_{t=1}^{T}$ are multiplicatively distorted by a univariate[4] *iid*(0, 1) process $\{\omega_t\}_{t=1}^{T}$ and a new series of $\Delta\mathbf{x}_t^b$ is constructed using $\Delta\mathbf{x}_t^b = \hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\beta}}'\mathbf{x}_{t-1}^b +$ $+\sum_{i=1}^{k}\boldsymbol{\Gamma}_i\Delta\mathbf{x}_{t-1}^b + \boldsymbol{\epsilon}_t^b$, $t = 1, \ldots, T$, where $\boldsymbol{\epsilon}_t^b = \omega_t \cdot \boldsymbol{\epsilon}_t$ with $\{\omega_t\}_{t=1}^{T} \sim iid(0,1)$, $\Delta\mathbf{x}_0^b = (\mathbf{x}_0, \mathbf{x}_{-1}, \ldots, \mathbf{x}_{-k+1})'$.

A wild bootstrap *p*-value of a Johansen quasi-LR test with a null hypothesis of cointegration rank *r*, for *B* replications of wild bootstrap and sample length *T*, is calculated by: $\tilde{p}_{r,T} = B^{-1}\sum_{b=1}^{B} I(Q_{r,b} > Q_r)$, where *I* is an indicator function, $Q_{r,\,b}$ is a quasi-LR test value calculated for a VECM model estimated using series $\Delta\mathbf{x}_t^b$ constructed in a *b*-th replication of the wild bootstrap procedure, and Q_r is a quasi-LR test value calculated for a VECM model estimated using the genuine series $\Delta\mathbf{x}_t$.

Simulations (Cavaliere, Rahbek & Taylor 2008, 2010) under the null hypothesis of no cointegration and MGARCH heteroscedasticity innovations or unconditional heteroscedasticity innovations, have shown a reduction in test size distortion for the presented wild bootstrap variant in comparison to tests using asymptotic critical values for quasi-LR Johansen rank tests. These bootstrap tests are associated with a VECM model that assumes linear error-correction and short-term dynamics.

Some cointegration tests assume in their alternative hypotheses models with a specific type of nonlinear error-correction and short-term dynamics, but according to simulations they suffer from unacceptably large test-size distortions under MGARCH heteroscedastic innovations (Maki 2013). It is of more benefit in the statistical arbitrage problem to use cointegration tests that do not require advance specification of the model dynamics.

The extended definitions of integrated and cointegrated processes presented earlier in this paper can be referred to the Breitung cointegration rank test (Breitung 2002), which is asymptotically free of the nuisance parameter of long-term covariance, influenced by short-term dynamics

[4] Most frequently for process variables we assume Rademacher, standard normal or some discrete asymmetric distribution with $E(\omega_t) = 0$ and $E(\omega_t^2) = 1$.

(linear/nonlinear, number of lags included) and by potential conditional heteroscedasticity and parameters related to them. The Breitung cointegration test can be conducted without advance specification of a model. This is a very important aspect because in the statistical arbitrage problem it is not known in advance which assets have related log-prices processes. Specifying log-price models for numerous subgroups from an adopted universe of assets would be problematic. Instead, subgroups of cointegrated log-prices need to be identified by automatic searching, and Breitung test *p*-values (with a null hypothesis of no cointegration) can be applied to measure the strength of the relationships. This is a combinatorial optimization problem, which can be solved using a genetic algorithm with binary coding of solutions (with 1 when the asset log-price is included in the relationship) and a fitness function defined, for example, as 1 – *p*-value of a test with a null hypothesis of no cointegration.

There follows a short discussion of the Breitung cointegration rank test.

Let $\mathbf{E}_T = \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}'_t$ and $\mathbf{F}_T = \sum_{t=1}^{T} \mathbf{X}_t \mathbf{X}'_t$, where $\mathbf{X}_t = \sum_{i=1}^{t} \mathbf{x}_i$. The Breitung cointegration test incorporates the solution of a generalized eigenvalue problem:

$$\left| \lambda \mathbf{F}_T - \mathbf{E}_T \right| = 0.$$

For eigenvalue λ_j ($j = 1, \ldots, n$) we have:

$$\lambda_j = \frac{\mathbf{v}_j' \mathbf{E}_T \mathbf{v}_j}{\mathbf{v}_j' \mathbf{F}_T \mathbf{v}_j},$$

so when $\mathbf{v}_j$ belongs to sp($\boldsymbol{\beta}_\perp$) we have[5]: $\mathbf{v}_j' \mathbf{E}_T \mathbf{v}_j = O_p\left(T^2\right)$, $\mathbf{v}_j' \mathbf{F}_T \mathbf{v}_j = O_p\left(T^4\right)$ and $\lambda_j = O_p(T^{-2})$. On the other hand, when $\mathbf{v}_j \in sp(\boldsymbol{\beta})$ then for $T \to \infty$: $T^2 \lambda_j \to \infty$.

The Breitung test considers hypothesis H_0: $n - r$ common stochastic trends (r cointegration rank) against H_1: $< n - r$ common stochastic trends ($> r$ cointegration rank) and employs statistic:

$$\Lambda_{n-r} = T^2 \sum_{j=1}^{n-r} \lambda_j,$$

where $\lambda_1 \leq \lambda_2 \leq \ \ldots \leq \lambda_n$ are eigenvalues from the solution of a generalized eigenproblem.

Under the null hypothesis the test statistic has an asymptotic distribution derived using FCLT, which is a trace of a specified functional of ($n - r$)-dimensional standard Wiener process defined on [0, 1]. This distribution

[5] Derivations can be found in Breitung (2002).

is free of the nuisance parameter of long-term covariance. Under the alternative hypothesis, test statistic tends asymptotically to infinity, which means that the test has a right-side critical area.

According to the results of simulations (Maki 2013), the use of the Breitung cointegration test is recommended for samples of finite length, when the innovations are characterized by MGARCH conditional heteroscedasticity and a null hypothesis assumes no cointegration (the Breitung test has minimal size distortion among considered tests).

It must not be forgotten that when conditional or unconditional heteroscedasticity of innovations exerts a strong influence, the cointegration results returned by the tests can be spurious.

5. Conclusion

Cointegration between the log-prices of related assets is a necessary, but not a sufficient condition for the statistical arbitrage opportunity to hold. Idiosyncratic shocks that cause deviations from the equilibrium relation also increase the dispersion of the mispricing process. In this way the autoregressive tendency of the mispricing process (whose values approximate the value of the statistical arbitrage portfolio over time) can be masked by inflated conditional variance. Future movements of the mispricing process can be hard to predict and also opposite to those suggested by the expected value conditional on the process past. Another difficulty in implementing a strategy of statistical arbitrage under heteroscedastic innovations is the increased chance (with respect to the critical level assumed in the test) of finding false log-price relations in many types of tests with a null hypothesis of no cointegration.

Bibliography

Balke, N. S. and Fomby, T. B. (1997) "Threshold Cointegration". *International Economic Review* 38(3): 627–45, https://doi.org/10.2307/2527284.

Breitung, J. (2002) "Nonparametric Tests for Unit Roots and Cointegration". *Journal of Econometrics* 108 (2): 343–63, https://doi.org/10.1016/s0304-4076(01)00139-7.

Burgess, A. N. (2000) *A Computational Methodology for Modelling the Dynamics of Statistical Arbitrage*. PhD dissertation. London: University of London.

Cavaliere, G., Rahbek, A. and Taylor, A. M. R. (2008) *Testing for Co-integration in Vector Autoregressions with Non-stationary Volatility*. CREATES Research Paper No. 2008–50, Copenhagen: University of Copenhagen.

Cavaliere, G., Rahbek, A. and Taylor, A. M. R. (2010) "Cointegration Rank Testing under Conditional Heteroskedasticity". *Econometric Theory* 26 (6): 1719–60, https://doi.org/10.1017/s0266466609990776.

Chan, N. H. (2011) *Time Series: Applications to Finance with R and S-Plus*. New York: John Wiley & Sons.
Davidson, J. (1994) *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford: Oxford University Press.
Jarrow, R., Teo, M., Tse, Y. K. and Warachka, M. (2012) "An Improved Test for Statistical Arbitrage". *Journal of Financial Markets* 15 (1): 47–80, https://doi.org/10.1016/j.finmar.2011.08.003.
Johansen, S. (1995) *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford–New York: Oxford University Press.
Maki, D. (2013) "The Influence of Heteroskedastic Variances on Cointegration Tests: A Comparison Using Monte Carlo Simulations". *Computational Statistics* 28 (1): 179–98, https://doi.org/10.1007/s00180-011-0293-x.
Swensen, A. R. (2006) "Bootstrap Algorithms for Testing and Determining the Cointegration Rank in VAR Models". *Econometrica* 74 (6): 1699–1714, https://doi.org/10.1111/j.1468-0262.2006.00723.x.

Abstract

Arbitraż statystyczny – ujęcie krytyczne

Rozpatrywana w ramach strategii arbitrażu statystycznego dynamika procesu odchyleń od równowagi (*mispricing process*) ma charakter autoregresyjnego procesu stacjonarnego. Proces ten reprezentuje w przybliżeniu wartość w czasie portfela z wagami odpowiadającymi elementom wektora kointegracyjnego dla procesów logarytmów cen powiązanych instrumentów. Strategia polega na zajmowaniu długich bądź krótkich pozycji na wspomnianym portfelu na podstawie prognoz dotyczących kształtowania się procesu odchyleń od równowagi. W artykule przeprowadzono na gruncie teoretycznym analizę dotyczącą testowania kointegracji w przypadku warunkowej heteroskedastyczności procesów innowacji. Testy kointegracji wykorzystywane są w procedurze poszukiwania powiązanych procesów logarytmów cen instrumentów, które będą tworzyć portfel arbitrażu statystycznego. W pracy rozważano także charakter dynamiki procesu odchyleń od równowagi, będącego liniową kombinacją (elementy wektora kointegracji są jej parametrami) powiązanych procesów logarytmów cen, dla których zakłada się, że są generowane przez klasę modeli (T)VECM-GARCH. Przy takich założeniach dotyczących modelu procesów stawianie precyzyjnych prognoz dotyczących dynamiki procesu odchyleń od równowagi na podstawie przeszłych realizacji jest utrudnione. Praca może być punktem wyjścia do analiz empirycznych dotyczących konstrukcji portfela arbitrażu statystycznego. Wykorzystując rozważania teoretyczne, wskazuje się problemy, które można napotkać w badaniach empirycznych dotyczących konstrukcji opartej na kointegracji strategii arbitrażu statystycznego oraz modelowania i prognozowania procesu odchyleń od równowagi w przypadku warunkowej heteroskedastyczności procesu innowacji.

Słowa kluczowe: arbitraż statystyczny, kointegracja, warunkowa heteroskedastyczność, VECM-MGARCH, test kointegracji Breitunga.

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 15 • 2016
ISSN 1642-168X
AOC, 2016; 15: 91–107
DOI: 10.15678/AOC.2016.1506

Grażyna Trzpiot
Justyna Majewska

MODELLING LONGEVITY RISK IN THE CONTEXT OF CENTRAL STATISTICAL OFFICE POPULATION PROJECTIONS FOR POLAND TO 2050

Abstract

The problem of an ageing population confronts the majority of advanced countries. This paper analyses the probability, which may be termed the probability of a sustainable pension, that a retired person will not face financial ruin before they die.

Keywords: longevity risk, life expectancy projections, contribution pension plan, propability of ruin.
JEL Classification: C130, C180.

1. Introduction

Demographic ageing is the result of people living longer as mortality rates fall. In the majority of countries, the length of time people are expected to live has increased by 25–30 years during the last century. Of the social, political, economic and regulatory challenges presented by constant improvements in longevity, the consequences for pensions have perhaps received the most publicity (Barrieu et al. 2012). If improvements in life expectancy could be predicted, and taken into account when planning retirement, they would have a negligible effect on retirement finances

Grażyna Trzpiot, University of Economics in Katowice, Department of Demography and Economic Statistics, Bogucicka 14, 40-226 Katowice, Poland, e-mail: grazyna.trzpiot@ue.katowice.pl

Justyna Majewska, University of Economics in Katowice, Department of Demography and Economic Statistics, Bogucicka 14, 40-226 Katowice, Poland, e-mail: justyna.majewska@ue.katowice.pl

(Antolin 2007). Unfortunately, gains in mortality and life expectancy are uncertain. In this regard, longevity risk is associated with the risk that future mortality and life expectancy will not be as expected (Antolin 2007).

Rising life expectancy increases the risk that people will outlive the financial resources they have set-aside for retirement. For an insurer or a pension scheme, improving mortality rates raises the risk that pay-outs will exceed forecasts. There are roughly two types of longevity risk. The first, non-systematic risk, arises from random fluctuations between individuals and can be mitigated by increasing the size of portfolios, while the second, systematic risk, affects all individuals in a non-random manner and cannot be diversified by pooling. People are likely to be more concerned about non-systematic risks, while insurers are likely to be more concerned about managing systematic risks.

The main purpose of this paper is to understand how uncertainty regarding life-expectancy outcomes affect the liabilities of defined-contribution private pension plans provided by employers. To do so, the paper first focuses on assessing the uncertainty surrounding future developments in life expectancy, that is, longevity risk. Secondly, it examines the impact that longevity risk could have on defined-contribution pension plans provided by employers. In this paper we investigate the effect of systematic longevity risk.

2. Global Demographic Change

Increased life expectancy is a worldwide phenomenon. Improvements in health and the related rise in life expectancy are among the most remarkable demographic changes of the past century.

There are two ways in which the population may age (Arltová, Langhamrová & Langhamrová 2013):

– relative ageing of the population caused by a fall in the birth rate and the consequent fall in the number of children in the population,

– absolute ageing caused by a fall in mortality; there are then greater numbers of older people in the population due to rising life expectancy.

Whether the population becomes younger or older depends on the nature of the age structure in the past, and on current birth rate and mortality. High mortality rates in the past meant that life expectancy at birth was shorter. Global life expectancy, which rose from approximately 30 years in 1900 to 65 years in 2000, more than doubled in the twentieth century; it is forecast that it will have risen to 81 by the end of the twenty-first. Over the second

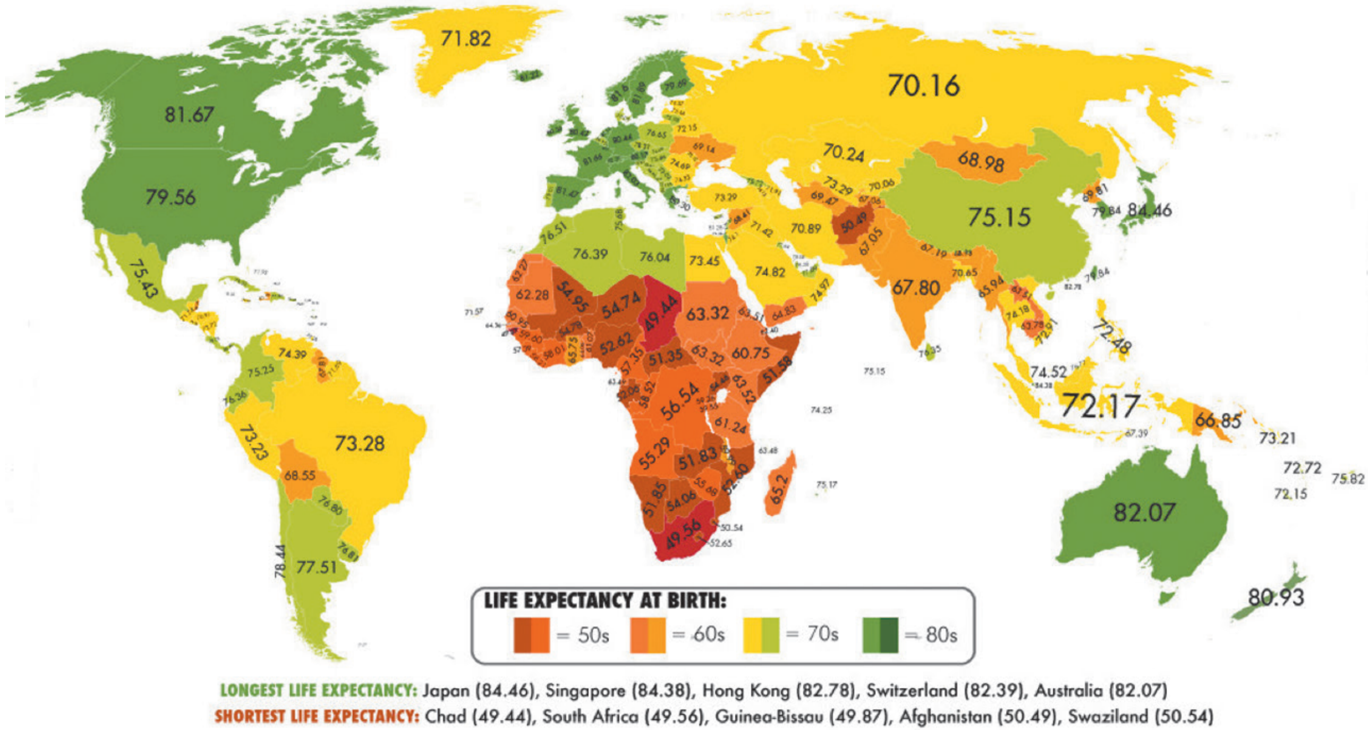

Fig. 1. Global Life Expectancy at Birth in 2013. Country Perspective

Source: CIA (2014).

half of the twentieth century, global life expectancy at birth increased by four-and-a-half months per year (2011), which amounts to a change of more than 18 years. The same upward trend is occurring in North America, South America, Europe, and Asia (Figure 1, Figure 2).

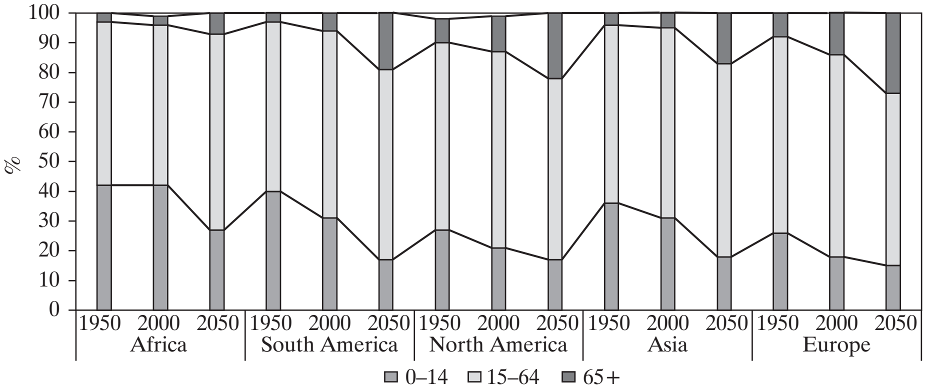

Fig. 2. Historical Trends and Projection of Age Group Shares in Selected Continent Populations

Source: www.un.org/en. Accessed: 10 March 2015.

Table 1. Longevity Trends (in Years), 1970–2050

Countries and regions	Observed			Projected	
	1970–2010	Increase per year	Standard deviation	2010–2050	Increase per year
Change in life expectancy at birth					
USA and Canada	8.2	0.20	0.14	4.3	0.11
Advanced Europe	8.6	0.21	0.13	4.7	0.12
Emerging Europe	1.1	0.03	0.36	6.8	0.17
Australia and New Zealand	10.8	0.27	0.27	4.9	0.12
Japan	10.8	0.27	0.23	4.6	0.11
Change in life expectancy at 60					
USA and Canada	4.9	0.12	0.11	3.1	0.08
Advanced Europe	5.7	0.14	0.13	3.7	0.09
Emerging Europe	0.6	0.02	0.18	3.8	0.09
Australia and New Zealand	7.2	0.18	0.23	3.7	0.09
Japan	7.7	0.19	0.19	3.7	0.09

Source: Human Mortality Database (13 December 2011) and IMF staff estimates.

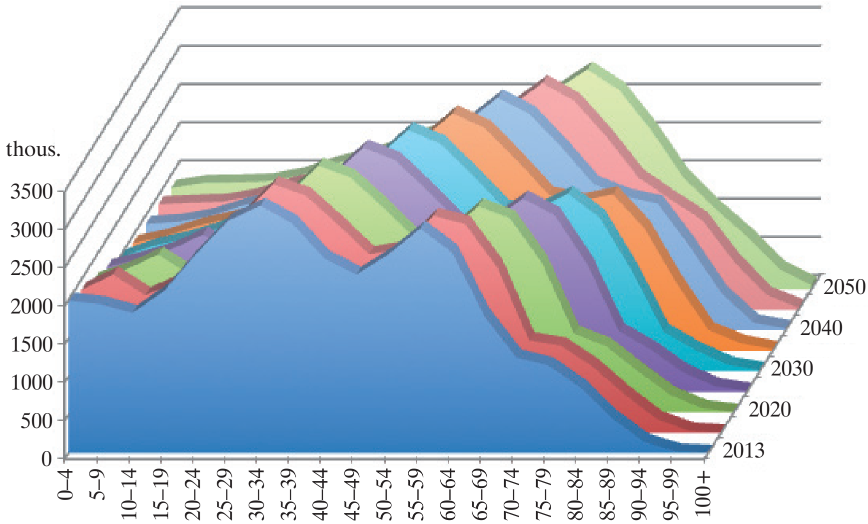

Fig. 3. Poland: Population by Age and Year

Source: population projection according to CSO in Poland, www.stat.gov.pl. Accessed: 10 March 2015.

The main source of longevity risk is the disparity between expected lifespans and actual lifespans, which has at times been considerable. Regardless of the technique used, forecasters have tended to consistently underestimate how long people will live (IMF 2012).

In 2009, numerous companies in developed economies closed their defined-benefit retirement plans. This represented a transfer of risk from industry and insurers back to policyholders. From a social point of view, this is no longer regarded as satisfactory. A number of countries, however, have been replacing defined-benefit pension plans with defined-contribution plans. But this has only resulted in the same unsatisfactory transfer of risk. Prompted by longevity improvements, ageing populations and the need to raise more finance for pensions, a number of governments are now planning to add an additional two to five years to the retirement age.

As Figure 3 illustrates, the changes that will occur in demographic age profiles will not leave Poland untouched. The average proportion of the population aged 60+ throughout our sample is projected to have increased to 29% in 2030 (compared to 16% in 1970), with most of the corresponding decline sustained by the group aged 0–19.

3. The Link between Mortality and Life Expectancy: Life Tables

In providing a summary description of mortality, survivorship, and life expectancy for a specified population, life tables represent a link between mortality and life expectancy. Complete life tables contain data for every single year of age, while abridged life tables contain data for five-year intervals and ten-year intervals. In its simplest form, a life table can be generated from a set of age-specific death rates (ASDR) which, based on vital statistics, are calculated as the ratio of the number of deaths during a year to the corresponding population size, which in turn is derived from censuses and annual estimates.

The final outcome of a life table is the mean number of years still to be lived by a person who has reached a specific age (hence age-specific life expectancies), if the current age-specific probabilities of dying are applied for the rest of their life.

In detail, this means that for each $x \in N$ up to a maximum age of, say, 120 (ignoring for the sake of clarity both truncated observations and cases of censored data, in which an individual's time of death is not precisely known), we consider the number l_x of individuals who turn age x. Assuming that d_x out of those l_x individuals will die between age x and $x + 1$, the annual mortality rate q_x at age x is the probability that someone aged x will die within one year. This can be estimated by l_x/d_x.

4. Longevity Risk

According to the NAIC definition (2010), this is the risk that actual survival rates and life expectancy will exceed expectations or pricing assumptions, resulting in a need for greater-than-anticipated cash flows for retirement. For individuals, this is the risk of outliving one's assets, which can lead to a lower standard of living, reduced care or a return to employment. For institutions that provide a guaranteed retirement income to people who are covered, longevity risk means underestimating survival rates. This results in increased liabilities and insufficient funds to make promised payments (NAIC 2010). The key drivers of the growing need to address longevity risk include an ageing population, increasing life expectancy, a shift in the locus of responsibility for providing a sufficient retirement income, the uncertainty of government benefits and economic volatility (NAIC 2010).

There are numerous holders of longevity risk. Principally they are governments, but they are also employers, individuals and insurers. There are various ways in which risk can be passed from one of these parties to another (Figure 4).

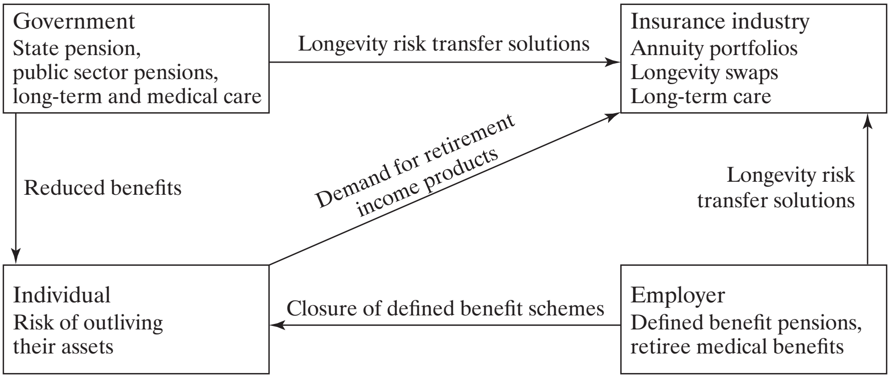

Fig. 4. The Holders of Longevity Risk

Source: Osorio (2013, p. 27).

There now follows an account of the description given by Swiss Re in 2014 of the relationships between holders of longevity risk. Given they undertake to pay retirees an income via a state pension, provide defined--benefit pensions for state employees and meet healthcare commitments, governments are influenced by an ageing society in many ways, all of which create significant liabilities. In an attempt to tackle this menace, many of them are beginning to reduce benefits in real terms, so that the burden placed on the individual to provide an income in retirement grows heavier. Employers who sponsor their employees' retirement incomes via defined-benefit plans, and employers who offer medical benefits to retired employees, will be concerned about the impact longevity can have on their future liabilities. To ameliorate this situation, many employers have closed plans down and replaced them with defined-contribution pensions, which has increased the risk burden on the individual still further. Given the declining amounts states and employers provide for retirement income, the responsibility placed on the individual is growing sharply. People are now expected to establish defined-contribution plans for their retirement and to address the risks associated with inflation, assets and longevity. There are therefore concerns that people will outlive the assets they have accumulated, which leaves a gap that the state is increasingly unable to fill. People are thus faced with the severe challenge of preparing for a stage in their lives when expenses, such as for long-term care, can be expected to increase. One solution is to work longer. But this depends on employment being available and on people being fit enough to do it. A well-diversified insurer will combine mortality risk, which is the risk that people will die sooner than expected, with longevity risk and other non-correlated insurance perils, such as property and casualty. It is this type of diversification, balancing two opposing risks, and diversifying across a portfolio of insurance perils, that in many cases makes insurers the natural home for longevity risk. (Re)insurers offer a range of solutions that can help governments, employers and individuals to pass on some or all of their longevity risk.

Like systematic risk, longevity risk is not diminished by diversification. In short, longevity risk is real, global, and non-diversifiable.

5. Modelling and Projecting Longevity

The close relationship between mortality and longevity modelling appears clear when we consider survival probability. Mathematically, life expectancy would appear to be the product of correlated mortality rates, which is supported by the following expression for the survival probability until date $t + u$ of a person aged x at time t (Barrieu et al. 2012):

$$S_t(x,T) = \prod_{i=0}^{T-1} [1 - q(x+i, t+i)]. \tag{1}$$

Mortality models are usually used for both mortality and longevity risks.

The literature contains several approaches to the projection of mortality rates (Wong-Fupuy & Haberman 2004). Public pension systems, or private pension funds, providing defined pension benefits, require mortality projections to determine the number of people who will be entitled to a pension.

The three main ways of modelling life expectancy are (1) a method based on underlying biomedical processes, (2) methods based on explanation that employ causal forecasting and econometric relationships and (3) methods of extrapolation that take historical mortality trends and project them forward. It is worth noting that these approaches are usually combined.

Models based on extrapolation are the ones that actuaries, official organizations and national statistical offices use most often. They employ past data to express age-specific mortality as a function of calendar time and, as such, can be deterministic or stochastic (Antolin 2007). The main difference between these models is that deterministic models do not take uncertainty of life expectancy into account, which means that they are not equipped with standard errors or projection probabilities. The literature distinguishes extrapolative stochastic methods that are based on (1) the interdependent projection of age-specific mortality (including graduation models, CMI), (2) standard time series procedures such as the Lee-Carter method (Lee & Carter 1992), where a log-linear trend for age-specific mortality rates is often assumed for the time-dependent component and (3) econometric modelling, of which *P*-spline models offer an example (Antolin 2007).

National statistical offices tend, however, to extrapolate historical trends in a deterministic way, while actuaries use stochastic approaches that are more sophisticated. What is more, national statistical offices and actuaries use different populations for their mortality and life expectancy projections. From the mortality tables they produce, national statistical offices project life expectancy for the entire populations of their countries. But the mortality rates of participants in private pension plans can differ substantially from those of the overall population, which is why these plans use their own actuarial mortality tables. It is a well-known fact that mortality rates are lower, and life expectancy is higher, for women and for well-educated, high-income individuals (Goldman 2001, Drever, Whitehead & Roden 1996). The use of life tables differentiated by socio-economic group can, however, give rise to a different set of problems (Antolin 2007).

Table 2. Projected Life Expectancy in Poland until 2050

Year	A1		A2		A3	
	men	women	men	women	men	women
2013 (real data)	73.1	81.1	73.1	81.1	73.1	81.1
2015	73.5	81.5	73.6	81.5	73.7	81.6
2020	74.9	82.5	75.0	82.6	75.3	82.7
2025	76.3	83.6	76.6	83.8	77.2	84.0
2030	78.0	84.8	77.5	84.4	78.3	84.8
2035	39.1	85.6	78.5	85.2	79.6	85.7
2040	80.3	86.5	79.5	85.9	80.9	86.7
2045	81.6	87.4	80.6	86.7	82.4	87.8
2050	83.0	88.4	81.8	87.6	84.1	88.9

Source: *Population Projection 2014–2015* (2014).

In Poland, the majority of population forecasts (and therefore of life expectancy forecasts) are based on deterministic models and are calculated in low, medium and high variants of future development:

– medium variant (A1) – the "delay" of Polish mortality in relation to the developed countries will be maintained at the same level throughout the forecast period,

– low variant (A2) – the "delay" of Polish mortality will remain at the same level until 2025; thereafter the pace of reduction in mortality will slow down,

– high variant (A3) – the distance between Poland and the developed countries will gradually decline throughout the forecast period.

In each variant the demographic factors are estimated based on the extrapolation of actual values and include a number of preconditions for the development of the individual components of population development.

6. Capital Requirements and the Probability of Ruin

Future mortality and life expectancy should be estimated using a stochastic approach which, by attaching probabilities to different outcomes, makes it possible to assess uncertainty and risk. Future developments in mortality rates and life expectancy are uncertain, but some paths or trajectories are more likely than others (*Pension Fund…* 2010). Forecasts of mortality and life expectancy should therefore consider a range of the most likely outcomes and take account of the related probabilities. There is a trade-off between greater certainty and greater precision.

If a pension system is based on the fund principle we must decide how much to "save" annually during the accumulation phase and how much to "spend" annually during the decumulation (annuity) phase (Cipra 2010). In view of the many random aspects, the best approach is similar to that applied in modern finance: Value at Risk, whereby the highest loss that can occur with a given probability (tolerance) is calculated. In the context of pensions, this must be modified to the probability that the retired person will not be "ruined" before the moment of death (the probability of a sustainable pension). Its obverse is the probability of ruin (the probability of an unsustainable pension). This is closely connected with the practise of pension planning or of managing the risk of pensions (Cipra 2010). In terms of internal models, the Solvency II guidelines propose using Value at Risk to compute the capital required when an insurer prefers to develop its own framework for risk assessment (Barrieu et al. 2012). The methodology considered here is very different from that now in use in the banking industry[1].

Where defined-contribution plans are concerned, contributions are in most cases defined in advance as a percentage of a participant's salary. The pension should be sufficient to provide an adequate income for the rest of a participant's life, and possibly also that of a partner, and should remove the risk that participants will outlive their resources.

At the age of retirement, for example sixty-five, capital of w is accumulated in the participant's account, which will be decumulated by the corresponding annual pension payments (Cipra 2010). The pension plan is stochastic and supposes that benefits follow a geometric Brownian motion. In modern finance, the randomness of interest rates on the capital invested from the participant's account is usually modelled by geometric Brownian motion (Malliaris & Brock 1982). Here, capital S_t in time t can be evaluated beginning with capital S_0 (in time 0) as (Cipra 2010):

$$S_t = S_0 \cdot e^{B_t(\mu,\sigma)} = S_0 \cdot e^{\mu t + B_t \sigma}, \qquad (2)$$

where B_t is the classical Brownian process, μ is the drift modelling the trend of the capital investment, and σ is the volatility modelling the diffusion of the capital investment. Note that S_t has a log-normal distribution.

The second aspect of the randomness of pension plans we should consider is the future lifetime of an individual. The randomness of the future lifetime T_x of an individual aged x can be modelled in the simplest case by the exponential law of mortality (Cipra 2010):

$${}_t p_x = \exp\{-\int_x^{x+t} \lambda_x ds\} = \exp\{-\lambda_x t\}, \qquad (3)$$

where λ_x is the force of mortality at age x (that is, an infinitesimal version of the probability of death at the given age). Life expectancy at age x is:

$$e_x = E(T_x) = \frac{1}{\lambda_x}. \qquad (4)$$

A combination of models (2) and (3) produces the present value PV_x of the standard pension (where the unit of pay is an annual payment in continuous time) as random variable:

[1] The Value at Risk measure has been introduced to insurance only recently. It is therefore based on data for only one year. While in banking there is access to high frequency data, which allows daily risk measures to be calculated, Value at Risk is calculated by insurers for the whole year and is an assessment of solvency.

$$PV_x = \int_0^{T_x} \exp\{-(\mu t + \sigma B_t)\}\, dt. \tag{5}$$

The probability of ruin, or the probability of an unsustainable pension (Dufresne 1990, Milevsky 1997, 2006), is defined as (Cipra 2010):

$$P(PV_x > w) = P(\int_0^{T_x} \exp\{-(\mu t + \sigma B_t)\}\, dt > w), \tag{6}$$

where $w > 0$ is the sum in the participant's account at retirement age x, which can be approximated as:

$$P(PV_x > w) \sim \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \int_0^{1/w} z^{\alpha-1} \exp\{-(\frac{z}{\beta})\}\, dz = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \int_0^{w} y^{-(\alpha-1)} \exp\{-(\frac{1}{\beta y})\}\, dy, \tag{7}$$

where:

$\alpha = \dfrac{2\mu + 4\lambda_x}{\sigma^2 + \lambda_x}$, $\beta = \dfrac{\sigma^2 + \lambda_x}{2}$ and $\Gamma(\alpha)$ is the gamma function,

$\Gamma(\alpha) = \int_0^{\infty} z^{\alpha-1} e^{-z}\, dz.$

7. Simulation Analysis

The formulas presented in Section 2 enable us to perform the corresponding calculations for pension plans in Poland. In the simulation study we use:

1. Financial data: the technical interest rate can be used for the purpose of the investment formula (2).

2. Longevity data: we have used life tables for male and female in Poland in 2013. From the expected remaining lifetime (life expectancy) e_x at particular ages x given in these life tables it is easy to estimate the parameters λ_x according to formula (4).

3. Projections of life expectancy at age 65.

Projecting Life Expectancy at Age 65 for Poland

The Lee-Carter method, whose principle is relatively simple, is used to forecast life expectancy. It involves modelling age-specific mortality over time based on the following:

$$\ln(m_{x,t}) = \phi_x + \psi_x \gamma_t + \varepsilon_{x,t}; \quad x = 0, 1, ..., k-1; \; t = 1, 2, ..., T, \tag{8}$$

where $m_{x,t}$ are specific mortality rates at age x and in time t, constituting $k - 1xT$ by dimensional matrix **M** of specific mortality rates at age x and in time t, $e^{\varphi}x$ is the average profile of mortality at age x (irrespective of time t), ψ_x is the age-specific constant that represents the speed of fluctuation of mortality at a given age, as opposed to the total level of mortality γ_t in time t (γ_t can also be described as the total mortality index), and $\varepsilon_{x,t}$ is white noise. The identification model is ensured by conditions $\sum_{t=1}^{T} \gamma_t = 0$ and $\sum_{x}^{k-1} \psi_t = 0$. The construction of the forecast is based on the fact that parameters $\hat{\phi}_x$ and $\hat{\psi}_x$ are constant in time and the total mortality index, which is a one--dimensional time series, is modelled and forecast based on the Box-Jenkins methodology (Box & Jenkins 1970). ARIMA models are used to calculate the forecast. Then, using estimates of parameters $\hat{\phi}_x$ and $\hat{\psi}_x$, a forecast of age-specific mortality rates is obtained from the relationship of

$$\hat{m}_{x,t} = \exp\{\hat{\phi}_x + \hat{\psi}_x \hat{\gamma}_t\} \quad x = 65; \; t = 2015, 2020, 2025, 2030. \tag{9}$$

The results are presented in Table 3.

Table 3. Projected Life Expectancy at Age 65

e_x	2015	2020	2025	2030
Male	15.75	16.03	16.69	17.25
Female	19.99	20.58	21.43	22.61

Source: authors' own calculations in the R programming language.

The Probability of Ruin

According to formula (6), we have calculated the probability of ruin (the probability of an unsustainable pension) for a retirement age of 65 depending on a spending rate of $1/w$. In this way, a spending rate of 0.06 would mean, for example, that a pension account of PLN 500,000 would pay PLN 30,000 annually and PLN 2,500 monthly. Table 4 and Table 5 present results for a retirement age of 65 only. The calculations are performed separately for males and females, and for various values of investment drifts and volatilities: $\mu = 1\%$ and $\sigma = 5\%$; $\mu = 2.25\%$ and $\sigma = 5\%$; and $\mu = 5\%$ and $\sigma = 10\%$.

Table 4. Probability of Ruin for Male (i.e. Probability of an Unsustainable Pension in %) for Retirement Ages and Spending Rates for Different Strategies

Year of projection	w									
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Strategy no. 1 μ = 1%, σ = 5%										
2015	0.4	2.6	7.4	14.3	22.7	34.7	41.1	49.9	58	65.2
2020	0.4	3.0	8.4	16.0	25.2	35.9	44.5	53.5	61.7	68.8
2025	0.5	3.1	8.5	16.3	25.6	37.4	45.0	54.1	62.3	69.4
2030	0.4	2.9	8.0	21.7	32.9	44.2	54.7	64.0	71.9	78.4
Strategy no. 2 μ = 2.25%, σ = 5%										
2015	0.1	1.2	4.0	8.8	15.3	23.0	31.4	39.9	48.1	55.9
2020	0.1	1.2	4.2	9.0	15.6	23.4	31.8	40.4	48.7	56.5
2025	0.2	1.3	4.6	9.9	17.1	25.4	34.3	43.4	51.8	59.7
2030	0.4	1.1	6.1	13.0	21.9	32.0	42.2	52.0	61.0	68.7
Strategy no. 3 μ = 5%, σ = 10%										
2015	0	0.4	1.7	4.2	8.0	13.0	19.0	25.6	32.7	39.8
2020	0	0.4	1.7	4.3	8.2	13.3	19.5	26.2	33.3	40.5
2025	0	0.5	2.0	4.8	9.1	14.7	21.4	28.7	36.3	43.8
2030	0	0.4	2.3	5.8	11.0	16.7	25.2	33.3	41.7	49.7

Source: authors' own calculations in the R programming language.

Table 5. Probability of Ruin for Female (i.e. Probability of an Unsustainable Pension in %) for Retirement Ages and Spending Rates for Different Strategies

Year of projection	w									
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Strategy no. 1 μ = 1%, σ = 5%										
2015	0.5	4.2	11.3	21.0	32.0	43.2	53.6	62.9	70.8	77.4
2020	0.2	4.4	11.7	21.7	32.9	44.2	54.7	64	71.9	78.9
2025	0.6	4.5	12.1	22.4	33.8	45.2	55.8	65.1	72.9	79.3
2030	0.7	4.8	12.8	23.6	35.4	47.1	57.7	66.9	74.6	80.8
Strategy no. 2 μ = 2.25%, σ = 5%										
2015	0.2	1.7	5.9	12.6	21.4	31.2	41.3	51.0	59.9	67.7
2020	0.1	1.9	6.1	13.0	21.9	32.0	42.3	52.0	61.0	68.7
2025	0.2	1.9	6.5	13.9	23.3	33.7	44.3	54.3	63.2	70.9
2030	0.2	2.0	6.6	14.1	23.6	34.1	44.8	54.7	63.7	71.4

Table 5 cnt'd

Year of projection	*w*									
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Strategy no. 3 μ = 5%, σ = 10%										
2015	0.0	0.6	2.3	5.6	10.7	17.1	24.6	32.7	40.9	48.0
2020	0.0	0.6	2.3	5.7	10.9	17.5	25.1	33.3	41.7	49.7
2025	0.0	0.6	2.3	5.8	11.0	17.7	25.3	33.6	41.9	50.0
2030	0.0	0.6	2.5	6.1	11.7	18.6	26.6	35.1	43.7	51.9

Source: authors' own calculations in the R programming language.

The results provide the following very interesting conclusions. Under a conservative investment strategy with parameters $\mu = 1\%$ and $\sigma = 5\%$, the probability that a man from Poland with a retirement age of 65 and a spending rate of 0.06 (an annual PLN 30,000 from a pension account of PLN 500,000) will face an unsustainable pension is 34.7% in 2015 and 44.2% in 2030. The probability is higher for a female of the same age: 43.2% in 2015 and 47.1% in 2030. If the investment drift increases, the probability of ruin falls considerably: for $\mu = 5\%$ and $\sigma = 10\%$, for example, the probability of ruin for males is only 13.0% in 2015 and 16.7% in 2030, while that for females is 17.1% in 2015 and 18.6% in 2030. This means that in 2015 only one in ten males and one in five females is ruined before death.

8. Conclusions

Demographic ageing must be understood as presenting a new challenge to society. There are a number of issues to be confronted if it is to cope with double the number of senior citizens, not the least of which is the rearrangement of systems of social and health care. It is important to remember that Poland is gradually becoming a longevity society. Unfortunately, gains in mortality and life expectancy are uncertain. Longevity risk, for example, which is defined as the uncertainty surrounding future developments in mortality and life expectancy, presents the threat that people will outlive the funds available to support them in retirement.

With respect to the randomness of investment activities and longevity, the model presented in this paper makes it possible to investigate the probability of the unsustainability of pensions. It provides numerical confirmation that this probability decreases as the age of retirement increases (Trzpiot & Majewska 2016), that it decreases as the spending ratio decreases (in particular where there is an increasing pension account and there are decreasing annuity payments), that it decreases as investment drift increases, that it decreases as investment volatility decreases, and that it is always lower for males than for females.

Bibliography

Antolin, P. (2007) *Longevity Risk and Private Pensions*. OECD Working Papers on Insurance and Private Pensions No. 3, Paris: OECD, https://www.oecd.org/pensions/insurance/37977228.pdf. Accessed: 5 March 2015.

Arltová, M., Langhamrová, J. and Langhamrová J. (2013) "Development of Life Expectancy in the Czech Republic in the Years 1920–2010 with an Outlook to 2050". *Prague Economic Papers* 22(1): 125–43, https://doi.org/10.18267/j.pep.444.

Barrieu, P., Bensusan, H., Karoui, N. E., Hillairet, C., Loisel, S., Ravanelli, C. and Salhi, Y. (2012) "Understanding, Modelling and Managing Longevity Risk: Key Issues and Main Challenges". *Scandinavian Actuarial Journal* 3: 203–31, https://doi.org/10.1080/03461238.2010.511034.

Box, G. E. P. and Jenkins, G. M. (1970) *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.

CIA (2014) *Factbook*, https://www.cia.gov/library/publications/the-world-factbook/fields/2102.html. Accessed: 5 March 2015.

Cipra, T. (2010) *Financial and Insurance Formulas*. Heidelberg–Dordrecht–London–New York: Physica-Verlag, Springer.

Drever, F., Whitehead, M. and Roden, M. (1996) "Current Patterns and Trends in Male Mortality by Social Class (Based on Occupation)". *Population Trends* 86: 15–20.

Dufresne, D. (1990) "The Distribution of a Perpetuity with Applications to Risk Theory and Pension Funding". *Scandinavian Actuarial Journal* 1: 39–79, https://doi.org/10.1080/03461238.1990.10413872.

Goldman, N. (2001) "Social Inequalities in Health: Disentangling the Underlying Mechanisms" in M. Weinstein, A. Hermalin, and M. Stoto (eds) *Population Health and Aging: Strengthening the Dialogue between Epidemiology and Demography*. New York: New York Academy of Sciences, pp. 118–39.

Graziani, G. (2011) "Live Long and Prosper". *Benefits Canada*, December: 64–65.

IMF (2012) "The Financial Impact of Longevity Risk" in *Global Financial Stability Report*. Washington: World Economic and Financial Surveys, pp. 123–53.

Lee, R. D. and Carter, L. R. (1992) "Modeling and Forecasting U.S. Mortality". *Journal of the American Statistical Association* 87 (14): 659–71.

Malliaris, A. G. and Brock, W. A. (1982) *Stochastic Methods in Economics and Finance*. Amsterdam: North-Holland.

Milevsky, M. A. (1997) "The Present Value of a Stochastic Perpetuity and the Gamma Distribution". *Insurance: Mathematics and Economics* 20 (3): 243–50, https://doi.org/10.1016/s0167-6687(97)00013-9.

Milevsky, M. A. (2006) *The Calculus of Retirement Income*. Cambridge: Cambridge University Press.

NAIC (2010) *Longevity Risk*. National Association of Insurance Commissioners and the Center for Insurance Policy and Research, http://www.naic.org/cipr_topics/topic_longevity_risk.htm. Accessed: 10 March 2015.

Osorio, J. (2013) *Mortality Improvements and Longevity Risk*. Swiss Re, http://www.fasecolda.com/files/5313/8912/6058/05-006-fasecolda_2013_seminario_de_vida.pdf. Accessed: 10 March 2015.

Population Projection 2014–2015 (2014). Statistical Analyses and Studies. Warsaw: Central Statistical Office, www.stat.gov.pl. Accessed: 12 March 2015.

Pension Fund Risk Management: Financial and Actuarial Modeling (2010) eds M. Micocci, G. N. Gregoriou and G. B. Masala. Boca Raton: Chapman and Hall/CRC.

Swiss Re (2014) *A Short Guide to Longer Lives: Longevity Funding Issues and Potential Solutions*, http://bvzl.de/Newsletter/2014_12/Spotlight_02.pdf. Accessed: 10 March 2015.

Trzpiot, G. and Majewska, J. (2016) "Modeling Longevity Risk". *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach* 288, *Informatyka i Ekonometria* 2016 (5): 33–46.

Wong-Fupuy, C. and Haberman, S. (2004) "Projecting Mortality Trends: Recent Developments in the United Kingdom and the United States". *North American Actuarial Journal* 8 (2): 56–83 https://doi.org/10.1080/109202777.2004.10596137.

Abstract

Modelowanie ryzyka długowieczności w świetle prognozy dla Polski do 2050 roku opracowanej przez Główny Urząd Statystyczny

Starzenie się społeczeństwa jest zjawiskiem, z którym mierzą się wszystkie kraje wysoko rozwinięte. W artykule analizujemy prawdopodobieństwo (zwane prawdopodobieństwem trwałej emerytury) wyczerpania zgromadzonych środków finansowych w okresie emerytalnym.

Słowa kluczowe: ryzyko długowieczności, projekcje przeciętnego trwania życia, program emerytalny o określonej wysokości składek, prawdopodobieństwo ruiny.

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 15 • 2016
ISSN 1642-168X
AOC, 2016; 15: 109–131
DOI: 10.15678/AOC.2016.1507

Valerio Fino
Janusz Rosiek

R&D: ITALY AND POLAND COMPARED

Abstract

Innovation and R&D are primary policy factors when governments seek to address major societal challenges through state-level welfare initiatives, and are essential elements in balancing socio-economic development. It is for this reason that the European Union and the governments of its member states have launched a series of initiatives, such as the Lisbon strategy, the Europe 2020 Strategy, the Green Paper on Innovation, and the Action Plan for Innovation in Europe, to stimulate and support innovation and R&D. The article reviews the literature on the theory of innovation in terms of its impact on R&D and compares the development activity undertaken in this area in Poland and Italy. Underpinned by quantitative data and a SWOT analysis, it then presents a comparative analysis of these countries' research systems. The conclusion specifies policy initiatives that could be taken by Poland and Italy to encourage innovation-based R&D.

Keywords: research and development (R&D), research system, innovations, EU innovation policy, SWOT analysis.
JEL Classification: 030.

1. Introduction

As well as being an important process in modern societies, research and innovation is a primary factor for governments wishing to address major societal challenges through state-level welfare initiatives. But how does innovation arise? With a focus on both empirical and theoretical considerations, the paper seeks to understand how R&D policy networks are activated by analysing the Italian and Polish research systems and describing

Valerio Fino, Università degli Studi di Bari Aldo Moro, Piazza Umberto I, 1, 70121, Bari, Italy, e-mail: valeriofino@gmail.com

Janusz Rosiek, Cracow University of Economics, Department of Theory of Economics, Rakowicka 27, 31-510 Kraków, Poland, e-mail: rosiekj@uek.krakow.pl

the various actors involved. The methodology combines desk research with a comparative analysis of the research systems of the two countries concerned.

The paper undertakes a thorough review of the research structure of Italy and Poland and places special emphasis on the role their governments and institutions play in innovation and economic growth. As it provides information that supports decision-making for strategies of economic development, R&D is one of the central issues of performance evaluation. It is for this reason that the analysis highlights the strengths and weaknesses of the Italian and Polish research systems from the perspective of the results that they achieve.

2. The Theoretical Background of R&D in the EU: A Brief Review of Innovation Theory and Important EU Documents

Onak-Szczepanik (2007) states that innovation management involves the constant search for new scientific findings and new ideas for producing new or improved materials, products, equipment, services, processes or methods intended for market or for other use (see also *Rola polskiej…* 2004). It should not be forgotten when we consider innovation in the economy, that a country's technological progress rests on three main actors: science, industry and government (Onak-Szczepanik 2007).

Stryjek (2015) points out that state-level measures to stimulate innovation involve providing businesses with relevant, broadly defined systemic conditions. In other words, innovation policy refers to the actions of central, regional or local state authorities in support of the creation, diffusion and use of innovation.

As Przychodzień (2013) notes, state innovation policy entails the conscious and purposeful action of public authorities in support of innovation in the economy. Its chief goal is to promote innovation and thereby make the economy more competitive. If successful, this will improve the quality and standard of living, signal the transition to a knowledge-based economy and promote cooperation between all market players (Ciok & Dobrowolska-Kaniewska 2009). Insofar as they encourage companies to invest their resources in innovation, the decisions taken by central and local government do much to create a favourable climate for investment. The important issues in this respect include infrastructure, the quality of the wider institutional order, the stability of legal norms, the ease and transparency of establishing and running businesses, the quality of human capital, the efficiency of public administration, and political stability

(North 1990, Witkowska 1996, Benacek et al. 2000, Djankov et al. 2002). The major instruments of innovation policy are legal, financial, institutional, infrastructural, structural and commercial (Dobrowolska-Kaniewska 2008).

The determinants of innovation include internal factors, such as human capital, experience, skills and accumulated knowledge (Klich 2016), and environmental factors, which either stimulate or inhibit innovation at SMEs (Batjargal 2007):

– initiatives and government projects (Hadjimanolis 1999),
– financial resources (Zhu, Sarkis & Lai 2011),
– cooperation between universities and companies (Veugelers & Cassiman 2005),
– changes in the market/dynamics (Martinez-Fernandez, Hinojosa & Miranda 2010).

Economists and scientists in Western Europe began to address state policy for the development of science and technology in the early 1980s, when the first books on the subjects appeared (Braun 1980, *Industrial Policy…* 1981, Rothwell 1986, Rothwell & Zegveld 1981, Stoneman 1987). As Jasiński (2013) observes, even if standardized terminology has yet to be adopted, significant advances have now been made in the worldwide literature on the subject. It should be noted in this connection that many authors use the terms “scientific and technical policy” and “innovation policy”, interchangeably.

It will not surprise us to learn that there is no unanimity as to the understanding of the concepts themselves (Jasiński 2013). Edquist (1994), for example, selects separate categories of R&D and technology to write of innovation policy, while the account of Furman, Porter and Stern (2002) defines priority areas for the development of science and technology, determines the level of support for research activities from public sources, and specifies ways to protect intellectual property. Dodgson and Bessant (1996), for their part, distinguish between scientific policy and technology and innovation. The aim of the latter is to improve the ability of companies, industries and the economy to innovate, and to facilitate the transfer of innovations. With admirable brevity, the European Commission (*Creating…* 2006) describes its approach as support for the formation of innovation--friendly markets.

Gibbons (1994) understands innovation policy in terms of a two-phase policy for science, whose development and maturity will lead to a broad range of policies for technical innovation. The current emphasis is on innovation policy in the regions (EC 2006, OECD 2011).

Innovation policy began appearing in the documents of the European Union in the 1990s. The first steps were taken with the publication of the Green Paper on Innovation in 1995 and the issuing of the First Action Plan for Innovation in Europe in the following year (Gust-Bardon 2011). The Lisbon Strategy, which was adopted by the European Council in Lisbon in 2000, represented a milestone in the approach to innovation policy (Rossi 2007).

The Green Paper on Innovation became the platform for European innovation policy in the years that followed. It set out to identify the positive and negative factors influencing innovation in Europe and to formulate measures to increase the EU's innovative capacity. The challenges and issues are diagnosed in chapters II (The Challenges of Innovation), III (The Situation in Europe) and IV (Innovation in a Straight Jacket), while the remedies are set out in Chapter V (Routes of Actions). The remedies include (Gust-Bardon 2011):

– facilitating administrative procedures,
– promoting the benefits of innovation,
– improving the financing of innovation,
– encouraging innovation at SMEs,
– modernizing the innovation efforts of the public sector.

3. The Italian R&D System

In the countries of southern Europe, including Italy, Spain, Greece, Malta, Cyprus and Portugal, investment in research does not exceed 1%, universities are struggling, there are few graduates, production is specialized and centred on medium and low technology, welfare provision is uneven and social inequality persists. In short, it is an area that tends to diverge from the rest of Europe (Greco 2011).

Where the socialization of risk is concerned, the conservative/corporatist welfare model is characterized by a greater emphasis on families, intermediate associations and voluntary provision. The provision of services in countries regarded as operating this model is mainly based on the principle of subsidiarity, which means that the state intervenes only where the family's capacity to provide for its members is exhausted. Italy did not produce a social security system on the lines of the Beveridge plan, and thus did not emulate the progress made in the north west of Europe, because of a comparatively low level of industrial development, and owing to the economic difficulties of the post-war period. Furthermore, its political culture meant that there was no social democratic party, and no liberal party worthy of the name. The ideological base and political focus were both absent. The victory of fascism brought down the curtain on an attempt to build a fair and universal welfare state. Even after 1945, Italy was not witness to the compromise between the working class and middle class that had added impetus to the evolution of the welfare state in the Scandinavian countries and Great Britain[1].

Italy will achieve its R&D target – albeit an unambitious one – if the current trend continues. Italy is a moderate innovator whose performance improved steadily up to 2012 before declining slightly in 2013. It was in the latter year that, nevertheless, the country reached a level of innovation of 80% compared to the EU as a whole. In the context of the Europe 2020 strategy, Italy set an R&D target of 1.53%, which was well below the EU average and insufficient to keep pace with the ever-shifting frontiers of technology in some sectors of its economy. In the aftermath of the 2008–09 financial crisis, austerity measures in countries such as Italy, Greece and Spain had a negative effect on research and may have jeopardised future generations of researchers. In that it reduced the public resources available, curtailed recruitment of new research personnel and introduced a drive to streamline the public sector, the economic crisis that broke out in 2008 had a radical impact on the response to the challenge to innovate. Table 1 presents a SWOT analysis for Italy.

Table 1. SWOT Analysis of Italy

	Strengths	Weaknesses
Internal	– universities – scientific co-publications – collaboration between academia and industry – women researchers – mobility and international attractiveness	– low R&D intensity – innovation performance – low business R&D investment – R&D expenditure inequality – foreign investment
	Opportunities	**Threats**
External	– state aid – investment incentives – international collaboration – attracting foreign researchers	– economic crisis – very high public debt – "brain drain"

Source: authors' own elaboration.

[1] Italy had to postpone its democratization process until after the defeat of fascism. Thus it was not until 1948 that a truly democratic constitution appeared and civil and political rights, and social rights underpinned by welfare, were codified.

Let us address the strengths that arise from the SWOT analysis of Italy.

Universities. They are responsible for 31.3% of total, national R&D expenditure. This is greater than in the EU 27, where the average national proportion in 2009 was 23.68%. There are 89 universities in Italy, of which 54 are state universities (JOREP 2011a).

Scientific co-publications. Italy, which according to the SCImago Institution Rankings (SIR) publishes 3.4% of international scientific publications, always occupies fourth place among European countries. What is more, the country is the world leader in academic publications when the ratio of publications to researchers is taken into account. According to our calculations, which were based on OECD and SCImago data, 726 articles per thousand Italian researchers were published in 2010, compared to 550 per thousand in the UK and approximately 400 per thousand in France and Germany (RIO 2016).

Collaboration between academia and industry. Law 240/2010 established a legal framework for this cooperation, which is based on a memorandum of understanding. The vast majority of universities and postgraduate schools offer programmes that receive the joint input of academia and industry. Thanks to their autonomy, Italian universities are free to establish bilateral relations with the business sector.

Female researchers. In 2010, 20.1% of grade A academic staff in Italy were women, which compares well with 18.6% for the Innovation Union reference group and the EU average of 19.8% (Deloitte 2013). At the policy level, a memorandum of understanding on gender equality in the research profession is in operation between the Ministry of Education, Universities and Research and the Ministry for Equal Opportunities. The latter ministry has also taken an active role in two EU-funded projects promoting gender equality. The Italian regional authorities have implemented specific measures to support the participation of female students in scientific programmes at universities (mostly at bachelor level) and to support women's careers through scientific training schemes.

Mobility and international attractiveness. In Italy in 2010 the percentage of non-EU doctoral candidates as a percentage of all doctoral candidates was 6.2%. This compared with 5.3% among the Innovation Union reference group and an EU average of 20.0% (Deloitte 2013). For reasons including the continued development of programmes taught in English, a number of Italian universities attract a higher number of non-Italian students and/ or doctoral candidates. Indeed in some cases the proportion is as great as 30%. The Rita Levi Montalcini Programme, which is a national fellowship programme managed by the Ministry of Education, Universities and Research, promotes the internationalisation of Italian universities by enabling early-stage researchers working abroad to carry out research projects at an Italian university of their choice. Its purpose is to recruit outstanding post-doctoral researchers working abroad and give them the opportunity to submit a proposal for a temporary position in conjunction with a proposal for a research grant.

Let us now consider the weaknesses revealed by the SWOT analysis.

Low R&D intensity. Public funding for R&D has been decreasing as a percentage of GDP over the last eight years. Italy set an R&D intensity target of 1.53% in the context of the Europe 2020 strategy, which is well below the current EU average and thus exposes some sectors of the economy to the risk of falling well behind the ever-shifting frontier of technology. In 2000–11, R&D intensity in Italy increased by an annual average of 1.69%, and rose from 1.04% in 2000 to 1.25% in 2010 (Deloitte 2013). While public sector and private sector expenditure on R&D have both increased during the period, the rate of growth has been modest. The difference between Italy's R&D intensity and the EU average is mainly due to lower industrial R&D. In 2011, business R&D intensity in Italy was 0.68% compared to an EU average of 1.26%. At 0.53%, public sector R&D intensity in Italy is also lower than the EU average, which was 0.74% in 2011 (Deloitte 2013).

Innovation performance. Italy remains below the EU average and its relative position has not improved significantly over the past five years: the synthetic innovation index was at 0.314 in 2004 and at 0.354 in 2008 (JOREP 2011a). Italy, which according to the European Innovation Scoreboard (EIS) belongs to the group of "moderate innovators", made slow progress and registered a below-average annual EIS growth rate of 1.8 in 2008 compared to the EU average of 2.3 (JOREP 2011a).

Low business R&D investment. Business R&D investment, which in Italy has been traditionally low, is highly concentrated in large firms and has grown weaker due to the recession that followed the economic crisis. Italy's lower level of business R&D intensity is partly due to the structure of the economy, in which the share of high-tech industry in total value added by manufacturing is low, and partly to low R&D investment by Italian firms. Though Italy remains non-specialized in all high-technology sectors except chemicals, there are cases of scientific specialization, such as in pharmaceuticals, or of high concentrations of patents, such as in other machinery and electrical equipment. The difficulty presented by very low business investment in R&D is aggravated by the size of Italian firms, of which 95% are small or micro enterprises. The proportion of foreign--owned firms is low and remained unchanged over the period 2001–08 (Deloitte 2013).

R&D expenditure inequality. Industrial R&D expenditure, of which 73.9% takes place in the north of Italy and only 10% at industrial firms in the Mezzogiorno in the south of Italy, is traditionally concentrated at a geographical level (JOREP 2011a).

Foreign investment. The stock of foreign investment in Italy accounts for only 12% of GDP, which is far less than in other EU countries. The main barriers to entry to the Italian market are labour taxes, lack of labour, inflexibility, bureaucracy and high corporate taxes. As domestic venture capital is scarce, the Italian government attempts to encourage foreign investors to invest in Italian companies. With the exclusion of the defence industry, foreign investors are permitted to invest in the privatization of government owned companies.

We now turn to the opportunities. Though it has not outstripped the average of 1.8% for the EU 25, total state aid in Italy has displayed a slight upward trend of 1.1% in recent years. State R&D aid has been boosted by additional funding from large strategic programmes managed by MIUR, and by the Industria 2015 initiative, which was launched by the financial law of 2007. The introduction of these two instruments heralded a switch to a more top-down R&D policy. Italy has adopted a range of measures to train sufficient researchers to meet its R&D targets, to promote attractive employment conditions at public research institutions and to address gender and dual-career issues. With the ultimate aim of attracting researchers, the government has taken initiatives to stimulate the interest of students in the natural sciences and in technology. As the number of enrolments in science and technology disciplines has increased by an average of approximately 20% against the 2008 baseline in recent years, the measures have proved successful (Deloitte 2013).

Investment incentives. The incentives the Italian government offers foreign investors are mainly designed to boost the economies of the more depressed areas – especially those in southern Italy. To add impetus to the development of a variety of industries, the Ministry of Education, Universities and Research has established programmes in eleven fields of development. They are meant to ease cooperation between public and private researchers and venture capitalists, support the research and development of key technologies, strengthen industrial research activities, and promote innovative behaviour at SMEs.

International collaboration. A number of partnerships, especially with the United States, Great Britain, France and Germany, are close to consolidation. A number of bilateral agreements are being developed by MIUR and other Ministries with EU countries and non-EU countries, but the majority of collaboration agreements are arranged and executed by research institutions and universities, including with the participation of private research organizations. Italy's policy aims to support joint programming and participation in international activities, research infrastructures and agreements. The chief source of funding is MIUR. The Ministries of Health and of Economic Development also make significant contributions to funding. There are no intermediary funding agencies.

Attracting foreign researchers. The major programmes open to foreign researchers in Italy concern mobility grants, which are supported by the International Inter-university Cooperation Fund. Under the auspices of the "Return of the Brains" and "Brain Gain" programmes, funds are also available to pay for Italian scientists to return from abroad. The goal of the "Futuro in Ricerca" and "Montalcini Programme" initiatives, which are funded by FIRB, is to attract foreign researchers to work in Italian academic institutions. These have not yet become open programmes. Italian research programmes funded by FIRB and open to the participation of foreign researchers are largely directed at supporting collaboration between Italian and foreign researchers and at attracting the latter to work in Italy.

Finally, let us consider the threats to Italy. The first is the economic crisis, which has paralyzed economic growth and investment in innovation. The government is working to introduce new reforms to escape the crisis and restart growth. Industrial production plunged during the recession while exports soared which, because Italian firms are strongly committed to investments in new EU countries, can be identified as a concern for the productive system. However, because it can stimulate the development of innovative policies, the economic crisis can be regarded as presenting an opportunity.

Very high public debt. With a public debt/GDP ratio close to 130% and a particularly heavy amortization of debt, Italy remains exposed to sudden changes of mood in the financial markets. Extensive and long-term policies to reduce public debt are therefore a priority. The results obtained thanks to recent structural reforms need to be consolidated and further measures need to be introduced to promote growth and improve competitiveness. If they succeed, they will lay a foundation for healthy growth.

Brain drain. As a result of the economic crisis, many Italian researchers and academics have been forced to leave Italy in search of better job opportunities and higher salaries.

4. The Polish R&D System

Poland is part of a region that is not homogeneous. It is a region characterized by research expenditure that rarely exceeds 1%, by lack of investment in universities, and by specialization in goods and services with little value added. Though it is in Europe's poorest area, it is exerting its efforts to prepare the structures that will form the base of a vigorous knowledge economy (Greco 2011).

Since the end of the communist period, Poland has achieved a high rate of development and an improved quality of life. In fact, following the fall of communism in 1990 and the transition from the People's Republic of Poland to the Third Republic, welfare spending has risen. The democratic political system, under which people have expressed the view that the state should meet social needs, has led to the growth of the welfare state. Since its return to democracy, Poland has faithfully pursued a policy of liberalizing the economy, so that today it is an exemplary case of the transition from a centrally-planned economy to a market economy.

There have been two formative periods for innovation policy in Poland. The first, which stretched from the early 1990s until Poland joined the European Union in 2004, began when the country recovered its political and economic freedom. It is characteristic that, due to the state's greater involvement in the transition from a centrally-planned economy to a market economy, innovation policy was then significantly curtailed. This was particularly true of the period until 1995. The marked decreases in the share of new and modernized products in industrial production (from 5.3% in 1989 to 3.4% in 1992), in the share of high-tech products in industrial production (from 10.3% in 1989 to 7.8% in 1992) and in the number of Polish inventions patented abroad (from 190 in 1989 to 43 in 1996), were all evidence of the country's low level of engagement in the modernization process.

According to Jasiński (2013), this period is associated with a lack of proper strategy and coordination between government agencies, insufficient R&D investment, over-centralized economic policies, the absence of a regional approach to shaping innovation policy, too little support – including financial support – for business research, and an emphasis on science in general terms rather than on innovation and the shortcomings of its transfer and diffusion. Nevertheless, with membership of the European Union drawing near and the need to adapt to the Lisbon strategy, innovation policy began to gather momentum at the turn of the century. Innovation programmes were launched at this time, which formulated goals for innovation policy. However, the specific instruments required to implement them were not yet in place. The period following EU accession was a great deal more important for innovation. Poland was now required to conform with the strategy of making the EU the most rapidly growing economy in the world: an economy that, based on knowledge, would bolster social cohesion by generating more and better jobs.

Jasiński's thesis concerning state innovation policy in a period of transformation of the national economy in Poland (2013) was sharply criticized by a number of authors (see, for instance, Bal-Woźniak 2012, Czerniak 2013, Janasz & Kozioł 2011, Jasiński 2006, Klincewicz 2008, Marciniak 2010, Moszkowicz 2001, Okoń-Horodyńska 1998, Pomykalski 2001 and *Zarządzanie*… 2010). As follows from the results of previous studies (Jasiński 2004), technical progress in the Polish economy in 1990–2004 was sustained by the influence of (a) macroeconomic adjustment, (b) market forces and (c) the influx of foreign technical ideas rather than by the impact of policy formulation and technological research. It should be noted, however, that there has been a significant qualitative change in this policy in recent years due to reforms to scientific research and a significant inflow of EU funds.

In comparing the Polish experience with that of other countries (EC 2003, OECD 1997), Jasiński (2004) proposed the following classification for the policy tools applied to innovation in Poland:

1. Regulations, especially state prohibitions, orders, limits, norms and standards relating to environmental protection, competition, and to consumer and intellectual property, which form a set of boundary conditions for the various actors in the field of innovation.

2. Systemic instruments, or financial incentives enacted in laws, that are intended to encourage companies to innovate.

3. Programmes and government projects, including public procurement, announced for a given period and designed to accomplish specific objectives and policy tasks in that period. This applies especially to the programmes run by the Polish Agency for Enterprise Development (PARP) and the National Centre for Research and Development (NCRD).

4. Bridging instruments that support intermediary bodies in the innovation process. These may be understood as providing "support through

support" for what are known as the institutions of the business environment. These tools are sometimes referred to as organizational or institutional. Table 2 displays a SWOT analysis for Poland.

Table 2. SWOT Analysis of Poland

	Strengths	Weaknesses
Internal	– EU structural funds: Poland is the country that benefits most from the structural funds allocated by the European Union every seven years – liberal law – rapid privatization – human resources: employees gain professional skills through a solid level of training using professional techniques and by institutions of higher education and additional courses providing specialized skills – gradual improvement in R&D – new optimistic forecast for R&D expenditure – new strategic research programmes – relatively high rates of economic growth	– welfare – lack of an "open, excellent and attractive research system" – low R&D intensity – low contribution from the business sector
	Opportunities	Threats
External	– foreign investment – venture capital funds – rapid growth of the global market	– centralization of R&D centres – "brain drain"

Source: authors' own elaboration.

Let us address the strengths that emerge from the SWOT analysis of Poland.

EU structural funds. Poland is the country that has benefited most from the structural funds allocated by the European Union every seven years. In 2007–13, Poland received EUR 102 billion, and the country will receive EUR 106 billion in 2014–20 – despite the crisis that has caused a general reduction in the EU budget (Szymański 2013). These funds have enabled strong growth compared to other EU states, which has been expressed in increased infrastructure, lower unemployment, increased exports (especially of food) and increased employment. By making proper use of EU funds, keeping rates to a minimum and providing support to SMEs, Poland has managed to avoid the recession that has afflicted the rest of Europe.

Liberal law and the privatization of firms. The privatization of small and medium state-owned companies and the presence of a liberal law on establishing new firms have permitted rapid development of an aggressive private sector.

Human resources. Employees have acquired professional skills and techniques through solid, general training, including that provided by institutions of higher education and by specialized bodies. Employees in Poland have a high level of motivation when it comes to accepting new challenges. They are willing to improve their qualifications, to acquire new skills and to accept new responsibilities. They accomplish this either by working for modern, international companies in Poland or by working abroad. In this way they have access to the latest technologies and can familiarize themselves with highly-developed organizations, in which high standards of staff training feed into continuous innovation of production processes.

Gradual improvement in R&D. Poland has become a moderate innovator. Poland's innovation performance improved only marginally in 2006–13, while its performance relative to the EU, where innovation increased more rapidly, fell from 54% in 2007 to approximately 50% in 2013 (EC 2014). Poland thus sustained its position as a moderate innovator until 2011, before becoming a modest one in 2012.

New forecast for R&D expenditure. Following publication of the EU2020 Strategy by the European Commission in March 2010, the Ministry of Science and Higher Education prepared a new R&D expenditure forecast for the period until 2020. It is expected that GERD (Gross Domestic R&D Expenditure) will increase to 1.7% in relation to GDP, with half of that amount coming from private funds.

New strategic research programmes. Poland has been distinguished in the past by having the highest share of research expenditure not addressed to specific social and economic objectives in the EU. But this is expected to change in the 2014–20 financial perspective with the adoption by the government in 2014 of Sixteen National Smart Specialisations, which were identified in a high-level policy document as likely to focus R&D efforts.

Economic growth. Expressed in Euro, GERD grew by an annual average of almost 12% in 2002–12. Yet Polish GERD, which was at 0.90% of GDP in 2012, remained one of the lowest in the EU when compared with the EU average of 2.06% (RIO 2015).

Welfare is a relative weakness. Following the fall of communism in 1990 and the transition from the People's Republic of Poland to the Third Republic, welfare spending has risen[2]. Transition has therefore not brought a reduction in the welfare state that was built during the communist period. Indeed, the welfare state has grown even bigger, so that social expenditure now accounts for a much larger share of GDP than before the transition. Social expenditure in Poland accounts for approximately 20% of GDP, and has remained roughly stable over the past several decades[3].

Lack of an "open, excellent and attractive research system". This is primarily due to the limited number of innovative companies, to unsuccessful linkages and entrepreneurship efforts and to scarce intellectual assets (patents, licenses, trademarks and designs).

Low R&D intensity. In terms of R&D employment and the number of research establishments, Poland can be described as a relatively large research system. Yet measured by R&D expenditure as a percentage of GDP, the country has a very low R&D intensity: against the background of the Lisbon GERD target of 3% of GDP, the 2007 average for the EU 27 was 1.85%, and the level in Poland was 0.57% (JOREP 2011b).

Low contribution from the business sector. Another feature of the R&D system is the consistently low contribution made by the business sector. In Poland, BERD accounts for only 0.33% of GDP and private enterprise is not especially active in R&D (Erawatch 2014b). Of the 1,000 EU companies ranked, only four Polish companies featured on the 2013 Industrial R&D Investment Scoreboard. The picture is, however, not entirely bleak: business expenditure on R&D has been rising gradually in recent years.

We now turn to opportunities, the first of which is foreign investment. Based on its comparatively low labour costs, Poland offers a strategic entry point to external investors looking to exploit its unfettered access to most EU markets. The Polish economy depends heavily on foreign funding, to the point that about two-thirds of its exports, which together account for 45% of GDP, are generated by companies established with foreign capital. Foreign investment brings innovation and new solutions for growth. Poland is one of the most attractive locations for venture capital funds. Where companies hit by hard economic conditions are undervalued and need restructuring, and new enterprises come to VC funds better prepared, the environment is

[2] The term "welfare spending" is defined as the share of public and private spending devoted to the welfare state.

[3] All of the data presented are available at: http://stats.oecd.org/Index.aspx?DataSetCode=SOCX_AGG. Accessed: 20 May 2015.

favourable for investors. A substantial share of investment will come from VC funds, which will look especially to the technology sector for attractive start-up companies.

Rapid growth of the global market. Poland combines low labour costs, including wages, with high-quality production. This makes the country attractive to foreign companies and multinationals, who will be prepared to move their operations and to invest the capital saved on production and logistics costs. Given the rapid growth enjoyed by the companies that invest there, Poland is tightly connected with international economic growth. Poland's export markets are recording steady and consistent growth, which is why the country is sometimes described as the "China" of Europe.

Finally, let us consider the threats to Poland.

Centralization of R&D centres. Many international companies active in Poland prefer to keep their R&D activities at headquarters, which limits relationships with other stakeholders.

Brain drain. Migration to higher-paid jobs in Western Europe poses a minor, but increasing, threat to the availability of skilled labour in Poland.

5. Conclusions

As can be seen from Table 3, the economic crisis that erupted in 2008 radically reduced the availability of public resources and limited the ability to hire new research personnel in Italy. The performance of the public sector was thus less effective and less efficient. With a population of 60 million, GDP per capita is 25,200 (compare and contrast with the target of 3% of GDP for R&D expenditure adopted in the Lisbon Strategy). Poland's R&D expenditure is among the lowest of the EU 28. Despite the respectable share of R&D in the public sector in that country, it is hard to innovate there. As Figure 1 shows, there has nevertheless been an increase over the previous year.

The Italian research system presents a number of difficulties. There would appear to be a poor attitude to implementation of results and cooperation with enterprises, which find it difficult to connect their research with input from the public research centres. Unlike in other countries, there may also be a certain resistance on the part of Italian public research bodies to adopting new organizational models and new incentive mechanisms. The Italian public research system is perhaps over populated. It has a large number of entities that activate relationships with foreign entities, which creates a fragmented R&D system and overlapping functions. There is also a degree of fragmentation in the sources of research funding. Both vertically at the level of the state, regions and sub-regions, and horizontally at the level of the ministries, regional councils, departments, universities and public research bodies, the Italian public research system is very articulated, each institution having specific responsibilities. The governance of the Italian public research system, understood as the procedures, and organizational and management tools, designed to integrate and coordinate the generation, dissemination and application of knowledge, is unstructured. It is therefore difficult to achieve the overall cohesion that would be conducive to accomplishing the country's strategic goals. The absence of institutional and programmatic unity also hampers relations with the European research system and restricts the ability to argue for the use of public and community resources for research and innovation.

Table 3. Comparison of R&D Policies between Italy and Poland. Year of reference: 2012

Italy	Poland
GDP PPP = 25,200 EUR R&D INT (GERD/GDP) (%) = 1.27 PR SEC SHARE OF R&D (%) = 55 PU SEC SHARE OF R&D (%) = 42 POPULATION (million) = 59.7	GDP PPP = 16,800 EUR R&D INT (GERD/GDP) (%) = 0.9 PR SEC SHARE OF R&D (%) = 37 PU SEC SHARE OF R&D (%) = 62 POPULATION (million) = 38.5

Note: GDP PPP – GDP per capita, R&D INT – R&D intensity, PR SEC – private sector, PU SEC – public sector.

Source: Erawatch (2014a, 2014b).

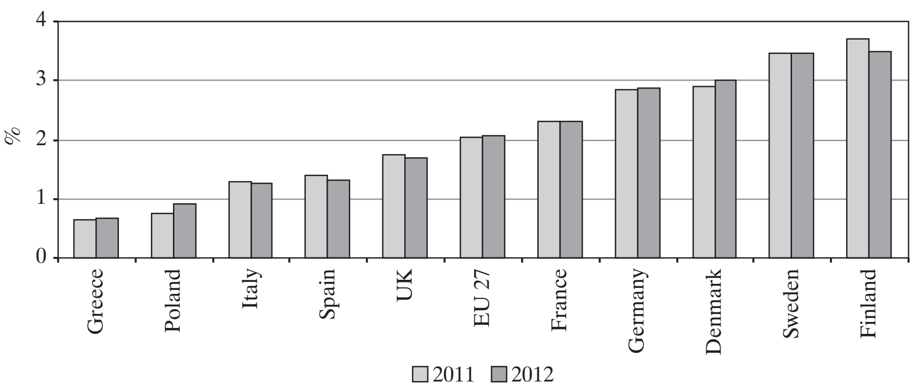

Fig. 1. R&D Intensity

Source: Eurostat Database, http://ec.europa.eu/eurostat/data/database. Accessed: 22 October 2015.

In that it has multiple agencies and research institutions that collaborate with the government, and a high organizational and thematic dispersion that results in a high proportion of research that is not focused on specified social and economic objectives, the Polish research system is different from the Italian one. It is a modern system that is in a transitory phase that will see significant changes in the roles of relevant actors and the introduction of multiple new rules. It continues to be driven by EU structural funds, which have made increases in R&D activities and in the number of R&D actors possible. The challenge now facing the government is to introduce new measures that will sustain the existing momentum and allow Poland to keep up with the major innovating countries in Europe. Poland has all of the assets required to make it an innovative country. It possesses one of the best European markets for foreign investment, it has a stable economy that is growing steadily, and there is a large internal market that affords easy access to neighbouring markets. Table 4 compares the challenges, differences, and similarities between the Italian and Polish research systems.

Table 4. Comparison between the Italian and Polish Research Systems

Factors	Italy	Poland
Challenges	Integrate and coordinate the processes of innovation	New innovative measures
Differences	Fragmented R&D system	High proportion of research not focused on specified social and economic goals
Similarities	Objective to spend 3% of GDP on R&D	Objective to spend 3% of GDP on R&D

Source: authors' own elaboration.

The research challenge facing Italy is to integrate and coordinate the various processes of innovation so that it can take its place as one of the most innovative countries in Europe. Given that the government's priorities lie elsewhere, and that the recession has made resources scarce, this is not a straightforward task. The problem that Poland must solve, meanwhile, is that of adopting new measures that will enable it to keep pace with other countries. If the government is able to turn the advantages bestowed by a robust economy in its favour, the country is well placed to make a success of introducing new measures. Turning to the differences, Italy presents a fragmented R&D system: there are a number of inefficient agencies that have failed to build the networks that would allow them to work together to achieve the required objectives. Though Poland has multiple agencies that are well connected to each other, it has a high proportion of research that is not focused on specified social and economic goals. The countries are similar in that they must both strive to bring their R&D expenditure to the level of 3% of GDP required by the European Union. The two countries share the mission of becoming more innovative and more competitive.

In recent years in Europe there has been a move away from traditional policy models, which have been gradually replaced by modern, third--generation innovation policies. Poland's version of this new approach has been to apply strategic planning to designated sectors and research strands, whose promotion, it is hoped, will lead to a knowledge-based economy. This movement has been stimulated by the provision of numerous new tools designed to support innovation and encourage scientific and business cooperation (Staśkiewicz 2013). Based on the analysis of innovation and R&D policy in Poland presented above, the following measures may be taken to shape its future character:

– introduce institutional reforms to implement innovation and R&D strategies,

– ensure adequate funding for the strategy to promote innovation (including tax incentives to stimulate innovative economic development),

– follow good practice for innovative solutions that support R&D activities,

– narrow the gap in meeting the needs of innovative development by improving the quality of scientific information, market information, and training and making them more easily available,

– interact more to increase the efficiency of academic centres and technology transfers, and to boost the public share in financing innovation in industry,

– create and develop infrastructure, organizations, and personnel that can support innovative entrepreneurship,

– promote the emergence of new companies based on high technology, especially in regions with high unemployment,

– increase the R&D efforts of foreign enterprises in Poland (within the framework of foreign direct investment),

– support the establishment and development of organizations that work at the crossroads of science and industry and form an important element in the infrastructure of technology transfer,

– aim to meet the criteria of good governance, especially in public consultation and information policy (openness and coherence),

– lighten the administrative burden, streamline the legislative process and enact economic law that is more consistent,

– create a system favourable to raising funds from internal sources; establish and develop institutions that will operate in the business environment,

– tailor academic education to the requirements of a modern, innovative, and competitive economy,

– devise and implement regional innovation policy while retaining focus at the national level,

– set an example to society in public administration by implementing innovative methods of cooperation and using new technologies and procedures in line with the recommendations of the European Union,

– promote innovative adaptations to export products and services so that they are in demand on world markets,

– increase public funding for research, education, and cooperation and move away from the reliance on grants as a source of capital, which is the dominant model in Poland,

– establish a stable legal framework for innovation policy,

– reduce the number of barriers to innovation,

– create scientific centres capable of gaining a good reputation in Europe and the rest of the world.

Bibliography

Bal-Woźniak, T. (2012) *Innowacyjność w ujęciu podmiotowym* [Innovation in a subject--based approach]. Warszawa: PWE.

Batjargal, B. (2007) “Comparative Social Capital: Networks of Entrepreneurs and Venture Capitalists in China and Russia”. *Management and Organization Review* 3 (3): 397–419, https://doi.org/10.1111/j.1740-8784.2007.00080.x.

Benacek, V., Gronicki, M., Holland, D. and Sass, M. (2000) “The Determinants and Impact of Foreign Direct Investment in Central and Eastern Europe: A Comparison of Survey and Econometric Evidence”. *Journal of United Nations* 9 (3): 163–212.

Braun, E. (1980) *Government Policies for the Stimulation of Technological Innovation*. Birmingham: University of Aston.

Ciok, S. and Dobrowolska-Kaniewska, H. (2009) *Polityka innowacyjna państwa a regionalny potencjał innowacyjny. Przykład Dolnego Śląska* [State innovation policy and regional innovation potential. The example of Lower Silesia]. Wrocław: Instytut Geografii i Rozwoju Regionalnego Uniwersytetu Wrocławskiego.

Creating an Innovative Europe (2006) ed. E. Aho. Brussels: European Commission.

Czerniak, J. (2013) *Polityka innowacyjna w Polsce* [Innovation policy in Poland]. Warszawa: Difin.

Deloitte (2013) *Researchers' Report 2013. Country Profile: Italy*. Deloitte.
Djankov, S., La Porta, R., Lopez-de-Silanes, F. and Shleifer, A. (2002) "The Regulation of Entry". *Quarterly Journal of Economics* 117 (1): 1–37, https://doi.org/10.1162/003355302753399436.
Dobrowolska-Kaniewska, H. (2008) *Potencjał innowacyjny i jego wykorzystanie w wybranych sektorach gospodarki Dolnego Śląska* [Innovation potential and its use in selected economic sectors in Lower Silesia]. Typescript of PhD thesis. Wrocław: Zakład Zagospodarowania Przestrzennego Instytutu Geografii i Rozwoju Regionalnego Uniwersytetu Wrocławskiego.
Dodgson, M. and Bessant, J. (1996) *Effective Innovation Policy: A New Approach*. London: International Thomson Business Press.
EC (2003) *Innovation Policy*. Brussels: European Commission.
EC (2006) *Science, Technology and Innovation in Europe*. Brussels: European Commission.
EC (2014) *Raffronto dei risultati degli Stati membri dell'UE, dei competitori internazionali e delle regioni europee nel campo dell'innovazione*, http://europa.eu/rapid/press-release_MEMO-14-140_it.htm. Accessed: 22 June 2015.
Edquist, C. (1994) "Technology Policy: The Interaction between Governments and Markets" in G. Aichholzer and G. Schienstock (eds) *Technology Policy*. Berlin–New York: W. de Gruyter.
Erawatch (2014a) *Erawatch Country Reports 2012: Italy*. Luxembourg: European Commission.
Erawatch (2014b) *Erawatch Country Reports 2012: Poland*. Luxembourg: European Commission.
Furman, J. L., Porter, M. E. and Stern, S., (2002) "The Determinants of National Innovation Capacity". *Research Policy* 31 (6): 899–933, https://doi.org/10.1016/s0048-7333(01)00152-4.
Gibbons, M. (1994) *The New Production of Science*. London: SAGE.
Greco, P. (2011) *La Germania cresce? Ha puntato su ricerca e welfare*, http://www.unita.it/scienza/politica/la-germania-cresce-ha-puntato-br-su-ricerca-e-welfare-1.293674. Accessed: 20 May 2015.
Gust-Bardon, N. I. (2011) "Polityka innowacyjna w Polsce" [Innovation policy in Poland]. Acta Universitatis Nicolai Copernici. Ekonomia XLII – *Humanities and Social Sciences*. *Nauki Humanistyczno-Społeczne* 403: 217–28.
Hadjimanolis, A. (1999) "Barriers to Innovation for SMEs in a Small Less Developed Country (Cyprus)". *Technovation* 19: 561–70, https://doi.org/10.1016/s0166-4972(99)00034-6.
Industrial Policy and Innovation (1981) ed. C. Carter. London: Heinemann.
Janasz, W. and Kozioł, K. (2011) *Innowacje w organizacji* [Innovation in the organization]. Warszawa: PWE.
Jasiński, A. H. (2004) "Czy polityka innowacyjna ma wpływ na innowacje?" [Does innovation policy impact on innovation?]. *Zagadnienia Naukoznawstwa* 1: 5–9.
Jasiński, A. H. (2006) *Innowacje i transfer techniki w procesie transformacji* [Innovation and technology transfer in the transformation process]. Warszawa: Difin.
Jasiński, A. H. (2013) "Instrumenty polityki innowacyjnej: czy grają w Polsce?" [Innovation policy instruments. Do they work in Poland?]. *Zagadnienia Naukoznawstwa* 1 (195): 3–23.

JOREP (2011a) *National Report on Joint and Open Programmes Italy*. Luxembourg: European Commission.
JOREP (2011b) *National Report on Joint and Open Programmes Poland*. Luxembourg: European Commission.
Klich, J. (2016) "Instrumenty pobudzania innowacyjności w sektorze MŚP" [Instruments for stimulating innovation in the SME sector]. Presentation during the panel discussion entitled "Stimulating entrepreneurial attitudes and innovation in Poland as a foundation for its future development" in a series of debates on "Development challenges for Poland in post-crisis conditions" organized by the Polish Economic Society, 4 May.
Klincewicz, K. (2008) *Polska innowacyjność: Analiza bibliometryczna* [Innovation in Poland. A bibliometric analysis]. Warszawa: Wydawnictwo Naukowe Wydziału Zarządzania Uniwersytetu Warszawskiego.
Marciniak, S. (2010) *Innowacyjność i konkurencyjność gospodarki* [Innovation and competitiveness]. Warszawa: C. H. Beck.
Martinez-Fernandez, C., Hinojosa, C. and Miranda, G. (2010) *Green Jobs and Skills: The Local Labour Market Implications of Addressing Climate Change*. OECD, CFE/LEED, www.oecd.org/dataoecd/54/43/44683169.pdf?conte. Accessed: 20 October 2015.
Moszkowicz, K. (2001) *Procesy innowacyjne w polskim przemyśle* [Innovation processes in Polish industry]. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego.
North, D. C. (1990) *Institutions, Institutional Change and Economic Performance*. New York: Cambridge University Press, http://dx.doi.org/10.1017/CBO9780511808678.
OECD (1997) *Choices of Policy Instruments*. Paris: OECD.
OECD (2011) *Regions and Innovation Policy*. Paris: OECD.
Okoń-Horodyńska, E. (1998) *Narodowy system innowacji w Polsce* [The national system of innovation in Poland]. Katowice: Wydawnictwo Uczelniane Akademii Ekonomicznej.
Onak-Szczepanik, B. (2007) "Innowacyjność gospodarki Polski a polityka innowacyjna państwa. Nierówności społeczne a wzrost gospodarczy" [The innovativeness of the Polish economy and state innovation policy. Social inequalities and economic growth] in M. G. Woźniak (ed.) *Gospodarka oparta na wiedzy* [The knowledge-based economy]. Working Paper No. 11. Rzeszów: Wydawnictwo Uniwersytetu Rzeszowskiego.
Pomykalski, A. (2001) *Innowacje* [Innovation]. Łódź: Łódź Polytechnic.
Przychodzień, J. (2013) "Realizacja polityki w zakresie wspierania innowacyjności w Polsce w okresie ostatniego kryzysu finansowego" [The implementation of policy to support innovation in Poland during the last financial crisis]. *Management and Business Administration. Central Europe* 21 (2(121)): 106–16, https://doi.org/10.7206/mba.ce.2084-3356.62.
Rola polskiej nauki we wzroście innowacyjności gospodarki [The role of Polish universities in fostering innovation in the economy] (2004) ed. E. Okoń-Horodyńska. Warszawa: PTE.
RIO (2015) *RIO Country Report 2014: Poland*. Luxembourg: European Commission.
RIO (2016) *RIO Country Report 2015: Italy*. Luxembourg: European Commission.
Rossi, F. (2007) *Innovation Policy in the European Union: Instruments and Objectives*. MPRA Paper (Munich Personal RePEc – Research Papers in Economics – Archive). Munich: Munich University Library.

Rothwell, R. (1986) *Public Innovation Policies*. Brighton: University of Sussex, SPRU.
Rothwell, R. and Zegveld, W. (1981) *Industrial Innovation and Public Policy*. London: F. Pinter.
Stanisławski, R. (2011) "Założenia polityki innowacyjnej państwa" [The goals of state innovation policy] in P. Niedzielski, R. Stanisławski, and E. Stawasz (eds) *Polityka innowacyjna państwa wobec sektora małych i średnich przedsiębiorstw w Polsce – analiza uwarunkowań i ocena realizacji* [State innovation policy vis-à-vis the SME sector in Poland – an analysis of conditions and an assessment of implementation]. *Zeszyty Naukowe* 654, *Ekonomiczne Problemy Usług* 70. Szczecin: Uniwersytet Szczeciński, pp. 13–36.
Staśkiewicz, J. (2013) "Polityka innowacyjna jako czynnik innowacyjności polskiej gospodarki" [Innovation policy as a factor in the innovativeness of the Polish economy]. *Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania* 33 (2): 173–187.
Stoneman, P. (1987) *The Economic Analysis of Technology Policy*. Oxford: Clarendon Press.
Stryjek, J. (2015) "Polityka innowacyjna i narodowy system innowacji w Polsce" [Innovation policy and the national innovation system in Poland]. *Kwartalnik Kolegium Ekonomiczno-Społecznego. Studia i Prace* 1: 59–76.
Szymański, L. (2013) *Budżet UE: ile i na co dostaliśmy*, http://gosc.pl/doc/1450783.Budzet-UE-ile-i-na-co-dostalismy. Accessed: 10 June 2015.
Veugelers, R. and Cassiman, B. (2005) "R&D Cooperation between Firms and Universities. Some Empirical Evidence from Belgian Manufacturing". *International Journal of Industrial Organization* 23 (5–6): 355–79, https://doi.org/10.1016/j.ijindorg.2005.01.008.
Witkowska, J. (1996) *Bezpośrednie inwestycje zagraniczne w Europie Środkowo-Wschodniej. Próba interpretacji na gruncie teorii bezpośrednich inwestycji zagranicznych i teorii integracji* [Foreign direct investment in Central and Eastern Europe. An attempted interpretation on the basis of foreign direct investment theory and integration theory]. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
Zarządzanie działalnością innowacyjną [Managing innovation] (2010) ed. L. Białoń. Warszawa: Poltext.
Zhu, Q., Sarkis, J. and Lai, K.-H. (2011) "Green Supply Chain Management Innovation Diffusion and Its Relationship to Organizational Improvement: An Ecological Modernization Perspective". *Journal of Engineering and Technology Management* 29 (1): 168–85, https://doi.org/10.1016/j.jengtecman.2011.09.012.

Abstract

Porównanie aktywności w sferze badawczo-rozwojowej w Polsce i we Włoszech

Innowacje oraz badania i rozwój (B+R) stanowią istotne sfery działalności w nowoczesnych społeczeństwach oraz zasadniczy element polityki rządów, które chcą sprostać kluczowym wyzwaniom społecznym związanym z implementacją koncepcji państwa dobrobytu do gospodarek narodowych. Ponadto aktywność w tych dziedzinach stanowi istotny element równoważenia ich rozwoju społeczno-gospodarczego. Z tego powodu Unia Europejska, a także rządy krajów członkowskich ugrupowania starają się stwarzać bodźce stymulujące rozwój tego rodzaju aktywności, przede wszystkim poprzez podejmowanie wielu inicjatyw mających na celu jej wspieranie, takich jak strategia lizbońska, strategia Europa 2020, zielona księga innowacji, plan działań na rzecz wspierania innowacji w Europie. W niniejszym artykule dokonano przeglądu literatury w zakresie teorii innowacji, w aspekcie jej oddziaływania na sferę B+R, a także porównano rozwój aktywności w tym obszarze pomiędzy Polską a Włochami. Podstawę prowadzonych rozważań stanowią: dane ilościowe, analiza SWOT oraz analiza porównawcza sposobu funkcjonowania systemów badawczych w tych krajach. W końcowej części opracowania zostały zamieszczone wnioski i rekomendacje odnośnie do pożądanego kierunku i charakteru polityki tych krajów w ramach wspierania działalności w zakresie B+R opartej na innowacjach.

Słowa kluczowe: badania i rozwój (B+R), system badawczy, innowacje, polityka innowacyjna UE, analiza SWOT.

REVIEWERS 2016

Prof. Sławomir I. Bukowski
Prof. Vesna Čančer
Prof. Oleksiy Druhov
Prof. Grzegorz Gorzelak
Prof. Janusz Korol
Prof. Bojan Krstić
Prof. Edward Nowak
Prof. Marek Ratajczak
Prof. Sergio Rossi
Dr Kalim Siddiqui

ISSN 1642-168X