ISSN 2059-190X (Online)

JOTA
Journal of Tax Administration

Volume 2, Issue 1 **April 2016**

Special Issue: The Shadow Economy

CONTENTS

ARTICLES

Reflections on the Meaning and Measurement of Unobserved Economies: What do we really know about the "Shadow Economy"? 5
Edgar L. Feige

Estimating the Underground Economy using MIMIC Models 41
Trevor Breusch

A Semantic Network Analysis of Laundering Drug Money 73
Martin Neumann & Nicholas Sartor

Drivers of Suspicious Transaction Reporting Levels: Evidence from a Legal and Economic Perspective 95
Julia Braun, Matthias Kasper , Alicja Majdanska, Maryte Somare

Behavioural Economics and Tax Evasion: Calibrating an Agent-based Econophysics Model with Experimental Tax Compliance Data 126
Cécile Bazart, Aurélie Bonein, Sascha Hokamp, and Götz Seibold

The Taxation of Micro and Small Businesses in Transition Economies: Country Experience of the Introduction of Special Tax Regimes 145
Michael Engelschalk & Jan Loeprick

UNIVERSITY OF EXETER | BUSINESS SCHOOL

CIOT
Chartered Institute of Taxation
Excellence in Taxation

Journal of Tax Administration

Editorial Board

Journal of Tax Administration

ABOUT THE JOURNAL

The Journal of Tax Administration is a peer-reviewed, open access journal concerned with all aspects of tax administration. Initiated in 2014, it is a joint venture between the University of Exeter and the Chartered Institute of Taxation.

JOTA provides an interdisciplinary forum for research on all aspects of tax administration. Research in this area is currently widely dispersed across a range of outlets making it difficult to keep abreast of. Tax administration can also be approached from a variety of perspectives including, but not limited to, accounting, economics, psychology, sociology and law. JOTA seeks to bring together these disparate perspectives within a single source, to engender more nuanced debate about this significant aspect of socio-economic relations. Submissions are welcome from both researchers and practitioners on tax compliance, tax authority organisation and functioning, comparative tax administration and global developments.

The editorial team welcomes a wide variety of methodological approaches including analytical modelling, archival, experimental, survey, qualitative and descriptive approaches. Submitted papers are subjected to a rigorous blind peer review process.

SUBMISSION OF PAPERS

In preparing papers for submission to the journal, authors are requested to bear in mind the diverse readership, which includes academics from a wide range of disciplinary backgrounds, tax policy makers and administrators and tax practitioners. Technical and methodological discussion should be tailored accordingly, and lengthy mathematical derivations, if any, should be located in appendices.

MESSAGE FROM THE CHARTERED INSTITUTE OF TAXATION

The Chartered Institute of Taxation is an education charity with a remit to advance public education in, and the promotion of, the study of the administration and practice of taxation. Although we are best known for the professional examinations for our members, we have also supported the academic study of taxation for many years and are pleased to widen that support with our involvement with this journal.

WEBPAGE

The Journal of Tax Administration website can be found here:
www.jota.website

ISSN 2059-190X (Online)

Editorial note

We are pleased to present this Special Issue of our journal, drawing on the papers presented and discussed at the 4th Shadow Conference that took place at the Tax Administration Research Centre, University of Exeter, on 23-25 July 2015. The themes of the conference were the shadow economy, tax evasion and fiscal intermediaries. We are grateful to all the authors that have contributed to this issue.

The first paper is one of the Keynote addresses of the conference: Reflections on the Meaning and Measurement of Unobserved Economies: What Do We Really Know about the "Shadow Economy"? by Edgar Feige. This paper discusses the various definitions of the shadow economy and the ways in which the different concepts can be measured. It also draws attention to some of the criticisms that have been made of the MIMIC macroeconomic estimation method that has been widely used to estimate the size of the shadow economy in a wide range of countries. The second paper, Estimating the Underground Economy using MIMIC Models by Trevor Breusch, was not presented at the conference, but was heavily cited by Professor Feige in his address and in discussions at the conference. The editors feel that it would help readers to be able to read this paper in its entirety. It provides a comprehensive analysis of the features of MIMIC models and raises concerns about the reliability of this methodology.

The third paper, A Semantic Network Analysis of Laundering Drug Money by Martin Neumann and Nicholas Sartor, describes a fascinating use of semantic analysis to analyse police records to determine the structure of a network that was involved in money laundering.The fourth paper, Drivers of Suspicious Reporting Levels: Evidence from a Legal and Economic Perspective by Julia Braun, Matthias Kaspar, Alicja Majdanska and Maryte Somare, is also about money laundering. It analyses data of the levels of reporting suspicious transactions in a variety of countries and uses both legal and statistical analysis to provide new insights into the characteristics of countries that influence the number of reported suspicious transactions. The fifth paper, Behavioural Economics and Tax Evasion – Calibrating an Agent-based Econophysics Model with Experimental Tax Compliance Data by Cécile Bazart, Aurélie Bonein, Sascha Hokamp, and Götz Seibold, describes a method of modelling tax evasion that exploits similarities between observed social interactions and the physical theory of magnetism. It calibrates the model and obtains results about the behaviour of individuals deciding whether or not to evade tax. The last paper, The Taxation of Micro and Small Businesses in Transition Economies: Country Experience with the Introduction of Special Tax Regimes by Michael Engelschalk and Jan Loeprick, makes use of the extensive experience of the authors in transition economies to describe and evaluate the many special tax regimes that transition economies have used, including the opportunities for tax avoidance that they sometimes provide.

Christopher Heady
On behalf of the Managing Editors

ISSN 2059-190X (Online)

Reflections on the Meaning and Measurement of Unobserved Economies: What do we really know about the "Shadow Economy"?[1]

Edgar L. Feige[2]

Abstract

This paper reviews the meaning and measurement of unobserved economies germane to tax evasion and macroeconomic information systems. These include unreported, non-observed, underground, illegal, informal and unrecorded economies. It reviews the progress and shortcomings of national and international agency efforts to measure these unobserved economies, noting what they have in common, what distinguishes one from another and their interconnections. It then examines the meaning of Professor Schneider's shadow economy (SSE), and the veracity of his claim to have accurately estimated its size and trend worldwide by employing a MIMIC model methodology. It concludes that SSE estimates suffer from conceptual flaws, apparent manipulation of results and insufficient documentation for replication, questioning their place in the academic, policy and popular literature.

INTRODUCTION

The past half century has seen an explosion of popular and professional literature referring to economies variously described as grey, black, subterranean, cash in hand, off the books, moonlight, undeclared, hidden, unofficial, concealed, parallel, invisible, occult, irregular, shadow, underground, non-observed, unreported, unrecorded, illegal, and informal. I propose that these economies comprise the activities of individuals, households and/or firms that evade, avoid, circumvent, elude, are excluded from, or not subject to the rules and conventions of established institutions. They all involve economic agents engaged in non-compliant behaviours that they seek to hide. This paper develops a broad conceptual framework, establishing what these various economies have in common, and what distinguishes one from another. It briefly recalls the discovery of non-compliant behaviours and reviews earlier empirical efforts to measure their magnitude. Its primary focus is on the meaning, measurement and consequences of non-compliance with fiscal codes, and violations of the rules and conventions of national income accounting.

[1] Keynote address delivered to the 4th Shadow Economy Conference, 24 July 2015 at the Tax Administration Research Centre (TARC), University of Exeter, UK. I wish to acknowledge the helpful comments of Joel Slemrod and Roberto Dell'Anno and the invaluable editorial help of my late friend and colleague, Allan Silver.

[2] Professor of Economics Emeritus, University of Wisconsin-Madison. Email: elfeige@wisc.edu.

The paper employs a national accounting framework to specify the meaning and conceptual interrelationships between unreported, non-observed, underground, illegal, informal and unrecorded economies. The literature often uses these terms interchangeably, although they have quite different meanings. We review the efforts of national revenue agencies to estimate the magnitude of tax evasion and the contributions of national and international statistical organizations to obtain exhaustive estimates of national income and output by measuring the non-observed economy. The paper's final section analyzes the meaning of Professor Schneider's shadow economy (SSE) and examines the veracity of his claim to have estimated its size, trend, causes and consequences worldwide (see, for example, Schneider, Buehn, & Montenegro, 2010a, 2010b, 2010c, 2011; Buehn & Schneider, 2012a; Schneider & Williams, 2013; Schneider & Enste, 2013).

Non-compliance

Since society penalizes non-compliant behaviours, entities that benefit from rule violations have incentives to undertake deceptive behaviours designed to defy detection, making non-compliant behaviours difficult and costly to observe and measure. Thus, non-compliance and "unobservability" are common features of all behaviour in unobserved economies (UEs). What distinguishes one UE from another is the particular rule being violated, making each UE distinctive in character, composition and magnitude. The impact of a particular non-compliant behaviour on the economy, society and polity depends on the nature and importance of the rule violated or avoided, and the extent of the rule violation.

Normatively speaking, the better the rule, the more harmful the consequences of non-compliance. Conversely, the worse the rule, the more beneficial the consequences of violating it. "Bad" rules typically inhibit voluntary exchanges, except where there are demonstrable negative external effects. "Good" rules prohibit coercive behaviours unless there are demonstrable overriding positive externalities (Epstein, 1995). Non-compliance with a bad law may be Pareto-improving, while breaking good laws may make society worse off (Leitzel, 1997). Non-compliant behaviours have real resource costs, as actors attempt to conceal their behaviours and authorities try to detect them. As well as affecting efficiency, non-compliance also has distributive consequences, shifting resources from the compliant to the non-compliant.

The study of non-compliant behaviours begins with the recognition that there are as many UEs as there are institutional domains with specific rules to be broken. For example, circumvention of the fiscal code by tax evasion gives rise to an *unreported* economy. Violations of the rules and conventions of national income accounting generate *unrecorded* and *non-observed* economies. Contravention of rules governing the production and distribution of prohibited goods and services, such as drugs, prostitution and human trafficking, gives rise to an *illegal* economy. Circumvention of

labour market regulations specifying minimum wages, working conditions, social security, unemployment and disability benefits gives rise to an *informal* economy that deprives some workers of deserved benefits while conveying undeserved benefits to others. Violations of the former Soviet Union's Five Year Plans, its production quotas and price controls, permitted a quasi-market reallocation of goods and services in what became known as the *second* or *parallel* economy (see Grossman, 1977; Feldbrugge, 1989; Ericson, 1984). Circumvention of immigration laws gives rise to an *illegal alien* economy; circumvention of currency exchange regulations to a *black market* economy; circumvention of intellectual property rights to a *knock-off* economy; circumvention of environmental regulations to a *pollution* economy, and circumvention of rules governing public officials' ethical behaviour to a *corruption* economy.

Given the variety of UEs, we limit our attention to those involving tax evasion resulting from violations of fiscal rules, and those affecting the nation's information systems due to circumventions of the rules and conventions of national income accounting. What kinds of observable traces does non-compliant behaviour leave behind that permits the social scientist to detect its presence? A relatively inexpensive means of hiding non-compliant transactions is to use cash as the medium of exchange. Since its usage does not leave a paper trail, it provides anonymity to those seeking to hide evidence of non-compliant behaviour. Anomalies in cash usage provide clues to identify, quantify and track changes in non-compliance over time.

Philip Cagan was the first to notice an empirical anomaly in currency usage during World War II (Cagan, 1958). Economists had predicted, and continue to predict, the advent of a "cashless society", anticipating a secular decline in the ratio of currency to deposits (or income) due to financial innovations and economic growth. Cagan's contradictory finding of an increase in the currency ratio led him to conclude that it was due to cash being used as a means to conceal income from the tax authority. He developed a simple currency ratio model that estimated US "unreported income" to be between nine and ten per cent of GDP ($21-$25 billion) in 1945.[3]

A second unexpected rise in the currency ratio during the 1970s and '80s suggested that tax evasion and unreported income might again be increasing. To the extent that national income and product accounts (NIPA) relied on income tax data as a basis for estimating components of aggregate output, there was concern that the national accounts might be systematically biased downward due to misreported income tax data resulting from tax evasion. A number of studies employing variants of Cagan's currency ratio method and Feige's (1979) transaction approach suggested that a relatively large and growing portion of the nation's economy had shifted from the

[3] Unreported income (a measure of the extent of fiscal non-compliance) is defined as the amount of net income not properly reported to the government due to non-filing, under-reporting income and/or overstating deductions credits and exemptions.

observed to the unobserved sector.[4] The "unobserved income hypothesis" (Feige, 1980, 1989a) maintained that observed "stagflation" could be partially explained by misguided macroeconomic policy based on biased estimates of income growth and unemployment due to an unnoticed shift from the observed to the unobserved sector of the economy.[5]

These academic findings stimulated the Internal Revenue Service (IRS) to estimate the extent to which growing non-compliance impeded the government's ability to raise revenues, and led the Bureau of Economic Analysis (BEA) to respond to concerns that the nation's national accounting information system might be systematically biased. The institutions whose rules were being violated had the greatest responsibility, incentives, resources and knowledge base to investigate these issues. They responded to the challenge with their own more detailed attempts to measure the extent and consequences of non-compliance.

The IRS undertook studies (1979, 1983) to improve audit strategies and to estimate the extent and nature of non-compliance with the US tax code.[6] Slemrod (2004, p.84) concludes that the IRS' Taxpayer Compliance Measurement Program (TCMP) produced "the most comprehensive, and probably most accurate, data on tax compliance for any country at any time". The TCMP consisted of 45,000-50,000 intensive "audits from hell" of sample tax returns for the years 1973, 1976, 1979, 1982, 1985 and 1988. Thereafter, the IRS abandoned the TCMP in response to strenuous political opposition. In 2001, the IRS initiated a new National Research Program (NRP) relying on less intrusive audits to estimate unreported income and the gross and net "tax gap" for the years 2001 and 2006.[7]

Suspicions that the nation's information system was biased motivated the BEA to clarify the relationship between unreported income on tax returns and unrecorded income that might be missing from the NIPA (Carson, 1984; Parker 1984). The BEA, well aware of deficiencies in the tax data on which it relies for measuring components of the national accounts, accordingly included "misreporting adjustments" in its estimates of income aggregates.[8] These considerable adjustments improve the accuracy

[4] Gutmann (1979) estimated what he called the "subterranean" economy, Feige (1979, 1980) the "irregular economy" and the "unobserved sector", and Tanzi (1983) and Feige (1989a) the "underground economy".

[5] Blinder and Rudd (2012) present evidence for a competing hypothesis, namely that the stagflation experience was entirely explicable in terms of supply shocks to the economy.

[6] See Feige (1989a, pp.33-36) for an account of the methods employed by the IRS.

[7] See IRS (2012); *Black* et al. (2012); Gemmell and Hasseldine (2012). The "gross tax gap", defined as the difference between the tax that taxpayers should pay and what is actually paid on a timely basis, was estimated at $345 billion in 2001. It rose to $450 billion in 2006. The "net tax gap" represents the amount of tax liability that will never be collected, despite enforcement efforts. This amounted to $290 billion in 2001 and $385 billion in 2006.

[8] Adjustments are made for non-farm proprietor income, corporate profits, interest paid, and wages and salaries on employment tax returns and not covered by unemployment insurance.

of NIPA accounts in the US, avoiding a key potential source of distortion.[9] The United Nations' System of National Accounts (ISWGNA, 1993) acknowledged the existence of underground, illegal and informal economies, but it was not until 2002 that a team of national accounts experts, drawn from national and international statistical organizations, produced *Measuring the Non-Observed Economy: A Handbook* (OECD, 2002).

Measuring the non-observed economy (NOE)

Spurred on by media reports of academic estimates, obtained by macro-model methods, of an alarmingly large unobserved economy (OECD, 2002, p.11),[10] the community of national income accountants collaborated in a major effort to obtain exhaustive and internationally comparable estimates of national income and product. This effort was particularly timely, given the disintegration of the Soviet Union and the expansion of the European Union. The countries of the former Soviet Union (FSU) and of Central and Eastern Europe (CEE) were replacing the central planning material product system (MPS) of national accounting with the United Nations System of National Accounts (SNA). The European Union required member states to adhere strictly to SNA accounting conventions, since it employed the resulting estimates of GDP to distribute grants and levy contributions. An added impetus came from a shift in economic production toward the developing world and the globalization of trade, requiring that the national accounts of developing nations accurately and exhaustively measure the level and growth of their productive capacities.

Consistent and exhaustive measures of GDP also form the basis for key policy decisions of international agencies. These often depend on ratio indicators, among them per-capita GDP, government debt and deficits to GDP, research and development expenditures to GDP and CO2 emissions to GDP. Recent estimates of VAT tax evasion also depend on the exhaustiveness of GDP accounting.[11] Recognizing that the quality of national accounts depends on the extent to which their coverage is exhaustive, the OECD's *Handbook* sought a common nomenclature among national statistical agencies and consistent methodologies representing what statisticians and national accounts experts regarded as best practice for measuring the NOE.

[9] In 2011, the last year for which all misreporting adjustments were tabulated, they totalled $1.3 trillion.
[10] The entire final chapter of the *Handbook* is devoted to a critique of these "macro-model methods".
[11] New "top down" approaches to measuring VAT tax gaps rely on national accounts aggregates. See Reckon (2009); Center for Social and Economic Research (2013a, 2013b); European Commission (2015); IMF (2013, 2014).

The national income accounting framework

National income accounting seeks to provide an exhaustive measure of productive economic activity. Let Y_e = exhaustive income defined as the sum of observed income (Y_o) and the income produced in the non-observed economy (Y_{NOE}):

$$1) \qquad Y_e \equiv Y_o + Y_{NOE}$$

According to the *Handbook*, the NOE comprises productive activities "that are missing from the basic data used to compile the national accounts because they are underground, illegal, informal, household production for own final use, or due to deficiencies in the basic data collection system" (OECD, 2002, p.3). If we include household production for own final use as part of the informal economy,[12] and deficiencies in the basic data collection system as reasonably included under the agency's definition of the underground economy,[13] the NOE (Y_{NOE}) is defined as the sum of the underground economy (Y_u), the illegal economy (Y_{il}) and the informal economy Y_{in}, such that:

$$2) \qquad Y_{NOE} = Y_u + Y_{il} + Y_{in}$$

The definition of the underground economy (Y_u) comes directly from the 1993 United Nations SNA, according to which the underground economy includes legal production activities that are "deliberately concealed from public authorities for the following kinds of reasons:

- to avoid payment of income, value added or other taxes;
- to avoid payment of social security contributions;
- to avoid having to meet certain legal standards, such as minimum wages, maximum hours, safety or health standards, etc.;
- to avoid complying with certain administrative procedures, such as completing statistical questionnaires or other administrative forms" (ISWGNA, 1993, p.153; OECD, 2002, p.139).

The illegal economy (Y_{il}) "consists of the income produced by those productive economic activities pursued in violation of legal statutes defining the scope of

[12] There is still considerable confusion concerning a precise definition of the informal economy, which varies depending on source. Household production for final use is not part of the informal economy according to the 15th ICLS Resolution (OECD, 2002, p.41); however, it is included in the Istat Analytical Framework (OECD, 2002, p.43).

[13] The *Handbook* refers to this as the "statistical" underground resulting from under-coverage, non-response or underreporting by enterprises (OECD, 2002, p.41).

legitimate forms of commerce" (Feige, 1990).[14] NIPA explicitly excludes illegal activities (Parker, 1984, p.19), since these are considered to be "bads" rather than "goods" and "because they are by their very nature conducted out of sight of public scrutiny and so data are not available to measure them" (BEA, 2009, Ch.2, p.2) However, both the SNA and the European System of National Accounts (ESA, 1996, p.61) require that all productive illegal activities be included in national accounts. Inclusion of the illegal economy is important to ensure that national accounts are consistent between countries and over time. Some activities, such as prostitution, may be legal in some countries and illegal in others. The legal status of some activities changes over time, for example the recent legalization of marijuana in some parts of the US. Only when national accounts are sufficiently exhaustive to include both legal and illegal production are they consistent between countries and over time (OECD, 2002, p.153).

Chapter 10 of the *Handbook* describes some of the complex definitional issues pertaining to informal sector production; however, there is no professional consensus concerning its definition. Charmes (2012, 2014) provides the most comprehensive coverage of the subject. For our purposes, it is sufficient to note that national accounts attempt to measure whatever non-observed production occurs in the informal sector.

Obtaining exhaustive measures of economic production requires measurement of the NOE (Y_{NOE}). Let β represent the fraction of the NOE that a statistical agency has succeeded in measuring and (Y^m_{NOE}) represent the amount of measured non-observed income. Then,

$$3) \qquad Y^m_{NOE} = \beta\, Y_{NOE.} = \beta(Y_u + Y_{il} + Y_{in})$$

Recorded income (Y_{rec}) – the published, official income aggregate – is the sum of observed income Y_o and measured non-observed income (Y^m_{NOE}). The "unrecorded" economy (Y_{urec}) is the difference between exhaustive income and recorded income[15]

$$4) \qquad Y_{urec} = Y_e - Y_{rec} = Y_e - (Y_o + Y^m_{NOE}) = Y_e - [Y_o + \beta\,(Y_u + Y_{il} + Y_{in})]$$

[14] The SNA defines illegal production as the "production of goods and services whose sale, distribution or possession is forbidden by law" and "productive activities which are usually legal but which become illegal when carried out by unauthorized producers" (ISWGNA, 1993, p.152). The OECD (2002, p.38) employs the same definition.

[15] Some fraction of "unreported income" due to tax evasion is recorded in measurement of the underground component of non-observed income in the form of a "misreporting adjustment". However, one must be very careful not to equate the underground component of NOE with "tax evasion", since unreported income from various sources such as capital gains does not reflect productive activities.

and $Y_{urec} \rightarrow 0$ as $\beta \rightarrow 1$.

Size of the non-observed economy

The *Handbook*'s publication enabled various national statistical agencies to undertake measurements of NOE, striving to produce exhaustive measures of GDP. The *Handbook* acknowledged that it is

> incumbent on national accountants to inform users of the extent of the *non-observed economy* – i.e., how much economic activity escapes direct measurement – and the extent of *non-measured economy*[16] – i.e. how much of the non-observed economy may still be missing from GDP after making the various adjustments of the kind described in this Handbook (OECD, 2002, p.192).

In 2003, the United Nations Economic Commission for Europe inventoried the practices of the 29 countries that had attempted to measure their NOE.[17] By 2008, this inventory included rudimentary descriptions of the estimation procedures employed by 43 countries and a sparse assortment of point estimates of their Y^m_{NOE} (UN, 2008, p.10). However, the UN surveys did not indicate how measures of NOE changed over time, nor did the surveys include information concerning the amount of measured NOE (Y^m_{NOE}) that each nation regularly included in its published national accounts statements.

National accounts measurements of NOE require a variety of imputations employing diverse statistical inferential methods to model lacunae in the basic data sources available to national accountants. Direct surveys and commodity flow approaches that balance supplies and uses of individual products contribute to the accounts' accuracy. Complex imputation methods vary from country to country, activity to activity and over time, requiring intensive documentation, reporting transparency, estimation of confidence intervals, and extensive professional oversight.[18] Adherence to these standards is necessary to mitigate misinterpretation of official statistics and to address concerns regarding the reliability of these exhaustive measures of GDP. To date, despite major expenditures of time and resources by national and international

[16] We have called this the "unrecorded" economy.
[17] The UN (2003, p.13, Table 1) erroneously lists the NOE adjustment of GDP for Kyrgyzstan in 1999 as 48 per cent. The correct figure is 13 per cent.
[18] Manski (2015) emphasizes the importance of communicating uncertainty in official statistics. Particularly germane is the issue of permanent statistical uncertainty arising from "incompleteness or inadequacy of data collection that does not diminish with time" (p. 637).

agencies, regular, comprehensive country reporting of the major components of recorded income, namely observed and measured non-observed income, is still lacking.

In the absence of any official compilation of measures of non-observed income since the 2008 UN survey, Table 1 presents time-series estimates of the ratio of measured non-observed income (Y^m_{NOE}) to GDP, generously provided by 27 statistical agencies from the Former Soviet Union (FSU) and Central and Eastern Europe (CEE) in response to my request for information.[19] Entries in bold print in Table 1 are those included in the latest UN publication (UN, 2008, p.10).

Whereas the statistical agencies of the US, Sweden, Netherlands and Australia report that measured NOE accounts for roughly one per cent of recorded GDP, the CEE and FSU countries' estimates range from five to 35 per cent and display considerable variation over time. The rising temporal pattern of some of the estimates may reflect a steep learning curve and the availability of improved statistical resources over time. Other apparent anomalies, such as the observation that Estonia's measured NOE is half the size of that of Lithuania's and Latvia's, or that Kazakhstan's is twice the size of Kyrgyzstan's, may be the result of respondents' reporting estimates being derived from different approaches to measurement. In theory, output, income and expenditure approaches to measurement should produce the same result; however, in practice they may differ considerably because they are derived from largely independent and less than perfect data sources. The UN's (2008) survey reveals that measured NOE estimates derived from the output approach are often twice as large as those estimated from the expenditure approach.[20] These discrepant results are an indicator of the degree of uncertainty associated with the estimation procedure, and typically serve as a signal that further work is required to reconcile the results.

The estimates reported in Table 1 all employ methodologies proposed by the *Handbook*, but the respondents did not identify their specific measurement approach, nor assess the reliability of their reported estimates, nor did they specify which of their estimates were finally included in their reported GNP statistics.

Exhaustive measures of economic aggregates are increasingly important in light of the growing extent to which investors and policy makers rely on national accounts data to guide consequential resource allocation decisions. Recent "top down" methods to estimate the VAT "tax gap" (Reckon, 2009; European Commission, 2015) rely critically on the exhaustiveness of the EU's national accounts to estimate the theoretical VAT tax liability which, when compared with actual VAT tax collections, yields the VAT tax gap estimate. The allocation of grants and levies similarly depends on the

[19] This is an updated version of Table 1 originally presented in Feige and Urban (2008, p.292).
[20] The UN (2008) reports that the Czech Republic, Latvia, Poland and Norway disclose both output- and expenditure-based estimates of measured NOE. The former are roughly twice the size of the latter.

exhaustiveness of EU member countries' national accounts. These dependencies make the accounts susceptible to "Campbell's Law":

> The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor (Campbell, 1975, p.35).

The recent inclusion of estimates of the size of the illegal economy in the recorded GDP of some European countries has heightened public skepticism regarding the reliability of the accounts.[21] If we are to gain greater confidence in their reliability, it is incumbent on international statistical organizations to monitor, assess and report on the accuracy of measures of the NOE that national statistical agencies include in their reported national accounts. All statistical agencies should be required to provide complete and timely reporting of measured NOE components included in reported GDP estimates. Gyomai and Van de Ven (2014, p.9) suggest that:

> Officially publishing estimates of the size of the NOE and its components … may help limit the proliferation of alternative estimates based on macro-econometric models, with the risk that these alternative measures eventually shape policies instead of the official national accounts embedded measures.

[21] According to *The Economist* (2015), illegal drugs and prostitution boosted the UK's GDP between 1997 and 2009 by £7 billion to £11 billion, accounting for roughly 0.7 per cent of GDP (ONS, 2014, p.4). Eleven other OECD member nations added illegal income to their GDP estimates. These ranged from 0.1 per cent of GDP for Germany to 0.7, 0.9 and one per cent for Poland, Spain and Italy respectively (Van de Ven, 2015, p.11).

Table 1. Measured non-observed income (YmNOE/(GDP)*100), 1990-2008

	Period Avg.	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05	'06	'07	'08
FSU																				
Armenia[1]	29.3					27.0	31.6	34.3	28.9	25.6	29.0	30.2	28.2	29.4	28.9	29.0	27.3	27.4		
Azerbaijan[2]	20.0											19.5	22.7	19.2	20.7	19.7	16.7	13.9		
Belarus[3]	10.4	5.7	6.0	10.7	11.0	10.1	11.9	13.2	10.9	11.0	11.7	11.1	10.6	11.1	10.7					
Estonia[4]	9.5				9.8	9.3	10.6	11.4	10.4	8.8	8.7	8.9	7.4	9.6	7.7					
Georgia[5]	30.7							26.9	27.4	30.3	30.3	33.7	33.4	33.2	33.1	28.3				
Kazakhstan[6]	25.2	12.2	13.0	21.7	23.0	24.3	32.9	37.1	37.9	30.3	27.4	24.7	23.9	22.6	21.6					
Kyrgyzstan[7]	12.7						8.4	9.4	10.3	11.9	13.2	13.1	14.4	16.5	17.0					
Latvia[8]	16.9								16.0	16.8		18.0	17.5	16.0	13.0	11.5	11.7	11.0		
Lithuania[9]	19.2			20.1	19.0	20.7	19.1	20.0	21.0	17.9	17.7	18.0	18.3	18.9		16.0	14.4	12.9		
Moldova[10]	30.5				33.0	29.6	26.2	24.2	31.4	30.1	34.4	34.6	31.6							
Russia[11]	13.6				5.3	8.5	10.4	11.7	11.9	11.9		24.8			24.3					
Tajikistan[25]	25.0														25.0					
Turkmenistan[25]	17.2													14.8			18.1			
Ukraine[12]	18.0											20.0	16.3	17.7	17.2	18.9				
Uzbekistan[13]	30.3									31.0			29.5							
CEE																				
Albania[14]	31.1							30.2	28.9	30.7	32.8	34.2	30.4	30.5	30.8					
Bulgaria[15]	18.3							27.8	31.2	12.3	12.0	16.3	10.2							
Croatia[16]	9.9						9.5	9.5	9.5	8.6	8.3	8.5	8.3	8.2	7.5	7.4	7.3			
Czech R.[17]	7.8									7.9	8.0	7.7	7.5	6.9	6.9	6.5	6.2	6.2	5.9	
Hungary[18]	15.6			16.0	16.0	16.0	16.0	16.0	16.0	16.0	16.0	11.9	16.0							
Macedonia[19]	14.7								14.4	16.2	13.7	12.9	14.9	14.4	16.3					
Montenegro[20]	6.6													7.8			5.3			
Poland[21]	15.4					17.2	16.6	15.9	15.2	15.3	14.5	14.6	14.3	15.4	15.8	14.5	15.9	15.9		
Romania[22]	16.5			6.7	8.9	12.6	16.6	18.4	18.6	23.3	21.1	21.1		17.7						
Serbia[25]	14.6														14.6					
Slovakia[23]	14.1						11.8	14.2	13.4	14.5	14.9	14.9	15.2	14.6	14.9	13.6	13.8	13.9	13.8	13.6
Slovenia[24]	7.1						6.4	6.3	6.5	6.1	5.8	6.9	6.8	7.5	8.0	8.0	8.0	8.0	8.0	

Sources: [1]Personal Correspondence (PC), Anahit Safyan, National Statistical Service, Republic of Armenia; [2]PC, Rasim Mirayev-SNA Division, Azerbaijan; [3]PC, Ministry of Statistics, Belarus; [4]PC, Andres Lauba, Statistical Office of Estonia; [5]PC, Revaz Tsakadze, State Department for Statistics of Georgia; [6]PC, Zhanara Isakova, Statbase, Kazakhstan; [7]Kudabaev (2004) & PC, Akylbek Masydykov, National Bank of Kyrgyzstan; [8]PC, Elita Kalnina, National Accounts Section, Latvia; [9]PC, G. Juskiene & I. Tvarijonaviciute, National Accounts Division, Statistics Lithuania; [10]PC, V. Gidilica, National Accounts Moldova; [11]Masakova (2001); [12]Golovko (n.d.); [13]Rogoznikova (2004); [14]Personal interviews at INSTAT, Tirana, Albania; [15]PC, T.Yalamov, Center for the Study of Democracy, Bulgaria; [16]PC, Robert Jurak, Central Bureau of Statistics, Croatia; [17]PC, N. Holikova & Vladimir Kermiet, National Accounts Department, Czech Republic; [18]PC, I. Bedekovics & Peter Szabó, National Accounts Department, Hungary; [19]UN (2003, p.225 & 2008, p.300); [20]PC, Branka Susic, Statistical Office of Montenegro; [21]PC, R. Popiński, Central Statistical Office of Poland; [22]Ciupagea (2004); [23]PC, V. Cicmanec, P.Baláž, & Jaroslav Sedivy: Statistical Office, Slovakia; [24]PC, A. Flajs, Statistical Office of the Republic of Slovenia; [25]UN (2008, p.10).

Macro-model estimation methods

The national accounting community concludes that the "lack of transparency in describing the procedures used to compile the national accounts is the main reason why outsiders resort to macro-models and produce estimates that undermine the credibility of the national accounts" (OECD, 2002, p.192). The *Handbook* devotes an entire chapter (Chapter 12) to criticizing these macro-model estimates – specifically, monetary methods, global indicator methods and latent variable methods – calling into question their relevance for national income accounting, the validity of their assumptions, and their stability, reliability and precision.

The limitations of currency demand methods in estimating the size of unobserved activities are well-known and extensively documented (for example, Feige, 1986, 1989a; Thomas, 1992; Bajada & Schneider, 2005; Breusch, 2005c). Small changes in assumptions regarding velocity, hoarding, dollarization, financial innovation, and benchmarking radically alter the resulting estimates, rendering them subject to wide error margins. However, to the extent that cash remains the preferred medium of exchange for transactions that individuals seek to hide, its temporal path contains clues to the evolution of non-compliant activities. The stubborn up-trend in per-capita cash holdings in the face of dramatic increases in currency substitutes remains the most significant trace evidence of increased unobserved activity over time (Feige, 2012a, 2012b).

The transactions method has not been used for four decades (Feige, 1979), given the proliferation of financial transactions and the difficulty of obtaining the data required to estimate their volume. Global indicator methods (e.g. electric consumption methods) have fallen out of favour (see Dobozi & Pohl, 1995; Kaufmann & Kaliberda, 1996; Johnson, Kaufmann, & Schleifer, 1997; Johnson, Kaufmann, & Zoido-Lobaton, 1998; Friedman, Johnson, Kaufmann, & Zoido-Lobaton, 2000; Eilat & Zinnes, 2002). Their simple assumptions are typically violated and the methodology produces anomalous results (Feige & Urban, 2003, 2008). However, latent variable methods have experienced a surprising resurgence since Frey and Weck-Hannemann (1984) introduced the innovation of treating the "hidden" economy as a latent variable.

Frey and Weck-Hannemann defined the hidden economy as "that part of the economy that escapes official measurement" and employed a multiple indicator multiple cause (MIMIC) model to estimate its size and trend in 17 OECD countries for the period 1960-1978.[22] They provided sufficient data and documentation to enable Helberger and Knepel (1988) to readily replicate the results and examine

[22] In the interest of full disclosure, Bruno Frey presented the original paper at the Netherlands Institute for Advanced Study 1982 International Conference on the Unobserved Economy. As editor of the conference volume, I rejected the paper (Feige, 1989a) because it was implausible to believe that the model's latent variable was measuring the "hidden" economy and because the most significant variable in the analysis (tax morale) had been arbitrarily constructed from a source found to be completely unreliable (see Strümpel, 1966). The paper was subsequently published as Frey and Weck-Hannemann (1984).

their robustness. While finding the methodology "basically meaningful and intellectually fascinating" (Helberger & Knepel, 1988, p.975), Helberger and Knepel concluded:

> A re-analysis of the data shows that the results of Frey and Weck-Hannemann's model are extremely unstable and cannot be regarded as reliable statements about the shadow economy of these 17 countries. Even minor variations in the number of countries included in the analysis, in the period under investigation or in the index which is defined, lead to a pronounced instability. A critical look at the observational variables used in the model lead to the conclusion that the authors cannot even be sure that what they have investigated is indeed the shadow economy. The latent variable could equally well be interpreted in other ways. It could be, for instance, that they have measured the development of the welfare state (Helberger and Knepel, 1988, p.965).

Giles (1999) and Giles and Tedds (2002) subsequently employed the MIMIC methodology to estimate the hidden economies of New Zealand and Canada. Their careful studies acknowledge the difficulty of interpreting the latent variable and the uncertainty associated with their resulting estimates. Hill (2000) and Smith (2002) criticized the complexity of Giles and Tedds' MIMIC methodology and, like Helberger and Knepel, found fault with the specification of causes and indicators. Breusch (2005a, p.388) severely criticized their MIMIC application, while commending the authors for documenting their calculations and providing their data in sufficient detail to permit replication and analysis of their results. Breusch's analysis finds that "radically different estimates are obtained when minor changes are made to the starting values of the estimation algorithm" (p.372) and that the temporal results were essentially derived from a single variable, interpretation of which could not be sustained as an index of underground activity relative to observed GDP. Moreover, the size of the underground economy was established using a benchmark from a separate, unidentified currency demand model, so the "estimates" of its key parameters are "merely numerical accidents without connection to the data" (Breusch, 2005a, p.387). Breusch concludes "that the overall level of the series estimated by Giles and Tedds is a mirage". Tedds and Giles' (2005) response to Breusch's critique acknowledges these shortcomings, admitting that they render "the specific results regarding the size and form of the underground economy reported in the book ineffectual" (Tedds & Giles, 2005, p.394).

Schneider's shadow economy (SSE)

Despite these critiques of MIMIC model applications, Friedrich Schneider and several co-authors subsequently employed the methodology to estimate what he calls the "shadow" economy.[23] Schneider's most recent claim – to have estimated the size and trend of the shadow economy for 162 countries to the accuracy of one decimal place[24] – has been widely cited in the academic literature and popular press, gaining visibility and influence through publication as a World Bank working paper (Schneider, Buehn, & Montenegro, 2010a, 2010b). What distinguishes Schneider's work from earlier MIMIC model applications (in addition to its ambitious scope) is the difficulty of obtaining documentation on his data, methods and sources to replicate his results and check their robustness and reliability.[25] Breusch's (2006a) review of Schneider and Bajada's (2005) study purporting to estimate the size of the shadow economy for 145 countries states the problem explicitly: "It is impossible to reconstruct these results from the documentation that is provided here or in other Schneider papers on which this chapter is based. Neither the data nor the model details were forthcoming from Schneider when I asked for them" (Breusch, 2006a, p.493).[26]

The problem of definition

The term "shadow economy" was originally introduced into the literature as the English translation of "Schattenwirtschaft", and is most often associated with the work of Professor Schneider, whose early papers used the term without defining it (Schneider & Neck, 1993). Schneider and Enste's (2000b) widely-cited survey paper on "shadow economies" mentions several possible definitions, but concludes: "In general, a precise definition seems quite difficult if not impossible" (Schneider & Enste, 2000b, pp.78-79). Among the *Handbook*'s key criticisms of "macro model results" is that:

[23] Schneider's MIMIC model specification for the OECD countries is virtually identical to the original Frey and Weck-Hannemann specification critiqued by Helberger and Knepel, except that it employs a more controversial "benchmarking" procedure. Compare Schneider and Williams (2013, p.47) with Frey and Weck-Hannemann (1984, p.40).

[24] The first mention of an error margin of MIMIC results appears *ex nihilo* in Schneider and Williams (2013, pp.30, 50): "Estimates of the size of the shadow economy by the MIMIC method are generally thought to have a margin of error of +/- 15 per cent". No explanation is offered as to how or where this error margin was derived.

[25] Breusch (2005b) succeeded in replicating Dell'Anno and Schneider's (2003) earlier Italian study and Bajada and Schneider's (2005) 17-country Asian-Pacific study.

[26] Over the past decade, I have encountered similar problems in attempting to obtain sufficient data and documentation to replicate Schneider's work (see Feige & Urban, 2008, p.288). Buehn and Montenegro have now provided the raw data for the Schneider, Buehn, and Montenegro (2010a, 2010b) study. However, requests for further documentation required for replication, concerning data sources, data inconsistencies, transformations and calibration specifications, have not been forthcoming.

> the activities that the models aim to measure are not precisely defined; it is often unclear whether the models are estimating non-observed or non-measured production, or whether they include informal or illegal activities as well as underground activities (OECD, 2002, p.187).

Possibly responding to this critique, Schneider subsequently adopted the SNA's 1993 definition of the "underground economy", Y_u (ISWGNA, 1993, p.153; OECD, 2002, p.139) as his own definition of the "shadow economy" (see, for example, Schneider, 2005, p.600, 2007, p.5, 2009, p.1081, 2010, p.443; Schneider & Buehn, 2009, p.2; Schneider, Buehn, & Montenegro, 2010a, p.5, 2010b, p.5, 2010c, p.444, 2011, p.55; Buehn & Schneider, 2012a, p.141, 2012b, p.175; Torgler, Schneider, & Schaltegger, 2010, p.305; Schneider & Williams, 2013, p.25). He explicitly excludes the illegal economy and the informal economy from his definition of the shadow economy. Schneider's shadow economy (SSE) therefore represents a single component of the NOE, namely the "underground economy" (Y_u), as defined by the SNA.

$$5) \qquad SSE \equiv Y_u$$

How, then, do MIMIC model estimates of SSE compare with estimates of measured non-observed income (Y^m_{NOE}) computed by national statistical agencies for comparable periods? Since SSE is defined to be only the underground component of non-observed income, it follows that, if β is close to unity, $SSE/Y^m_{NOE} < 1$. Table 2 reveals that, contrary to expectation, SSE (as a percentage of GDP) is on average two and a half times larger than the national accounting measures of the NOE for FSU and CEE countries. For the OECD countries for which we have comparable estimates, it averages eight times larger.

These discrepancies document why the national accounting community is so critical of macro-model estimates. The OECD's *Handbook* discusses macro-model estimates, "not because they are considered useful in obtaining exhaustive estimates of GDP or in estimating underground production, but because they tend to produce spectacularly high measures, which attract much attention from politicians and newspapers" (OECD, 2002, p.187). If Schneider's results were accurate estimates of the "underground" component of NOE, one would have to conclude that national accounting information systems have grossly failed to obtain exhaustive measures of national income, product and expenditure.

Table 2. Comparison of SSE with Y^m_{NOE} as a percentage of GDP

Country	SSE[1]	Y^m_{NOE} **	SSE/Y^m_{NOE}		Country	SSE *	Y^m_{NOE}[2]	SSE/ Y^m_{NOE}
FSU					**Other**			
Armenia	49.9	28.9	1.7		Australia	14.3	1.3	11.0
Azerbaijan	61.6	19.8	3.1		Austria	9.7	7.9	1.2
Belarus	50.2	11.0	4.5		Belgium	22.0	3.5	6.3
Estonia	39.3	8.7	4.5		Brazil	39.6	12.8	3.1
Georgia	68.1	32.0	2.1		Italy	27.0	15.8	1.7
Kazakhstan	45.2	24.0	1.9		Ireland	16.1	4.0	4.0
Kyrgyzstan	41.2	14.8	2.8		Mexico	30.5	12.1	2.5
Latvia	39.6	14.6	2.7		Netherlands[3]	13.7	1.0	13.7
Lithuania	30.5	17.2	1.8		Norway[3]	18.2	1.7	10.7
Moldova	47.3	33.5	1.4		Spain	22.7	11.2	2.0
Russia	48.9	24.6	2.0		Sweden	19.2	1.3	14.8
Ukraine	55.9	18.0	3.1		Turkey	31.0	1.7	18.7
Uzbekistan	37.2	29.5	1.3		United States	8.8	0.8	11.0
CEE								
Albania	35.1	31.7	1.1					
Bulgaria	37.4	12.8	2.9					
Croatia	35.7	8.6	4.1					
Czech R.	19.4	7.1	2.7					
Hungary	25.7	14.6	1.8					
Macedonia	38.3	16.3	2.3					
Poland	28.0	15.0	1.9					
Romania	36.5	20.0	1.8					
Serbia	41.1	14.6	2.8					
Slovakia	19.0	14.6	1.3					
Slovenia	28.0	7.3	3.8					

Sources: [1]Schneider & Williams (2013, pp.149-154); [2]Average values from Table 1 & UN (2008, p.12); [3]Schneider (2005, p.611).

Since the allocation of both public and private resources is heavily dependent on the accuracy of these national information systems, we must further examine the reliability and robustness of SSE estimates based on currency demand MIMIC model methods. To anticipate our conclusions, once the curtain of complexity surrounding the MIMIC procedure is lifted, we find that Schneider's estimates are so arbitrary, fragile and poorly documented that they cannot be taken seriously as estimates of the NOE, nor of tax evasion, as is sometimes erroneously claimed to be the case (see Schneider, 2012; Murphy, 2012, pp.11-12; Schneider, Raczkowski, & Mróz, 2015).

Examining the veracity of Schneider's mimic model estimates

Breusch's (2005a, 2005b) exemplary efforts to replicate and analyze the reliability of MIMIC methods sets a high professional standard rarely surpassed. In his detailed examination of studies by Giles and Tedds (2002), Dell'Anno and Schneider (2003), and Bajada and Schneider (2005), Breusch explains that "understanding their results requires peeling away the layers of econometric complication, which include MIMIC modeling, prediction, and benchmarking" (Breusch, 2005a, p.387). He carefully attempts to replicate the empirical results of each study, noting data transformations, calibration procedures and their consequences. He expresses particular concern about the "control that the researchers exercise over their methods to ensure that the results are interesting, and reasonable (meaning challenging but not too outlandish)" (Breusch, 2005a, p.388), demonstrates how key assumptions of the MIMIC model are violated (Breusch, 2005b, p.28), and skillfully demonstrates the arbitrary nature of the calibration procedures. He discovers:

> transformations of the data … are not documented … and as a result of these ancillary treatments, it is not always clear to the reader how, and by how much, the results of the MIMIC model are stretched and squeezed to fit some outside evidence. … The upshot is a method that lacks objectivity because it is open to manipulation and misrepresentation (Breusch, 2005b, p.3).

As we shall see, arbitrary choices made by the investigator determine the signs of the causal variables, the time path of the "shadow economy" and its size.

In order to assign a scale to the latent variable, the MIMIC model requires the choice of an indicator variable for normalization, typically given a unit coefficient (Bollen, 1989, Ch.6). Bajada and Schneider (2005) choose currency holdings as their normalizing indicator with a unit coefficient, while Dell'Anno and Schneider (2003) choose real GDP as their normalizing indicator, but specify that it must have a coefficient of *negative* one. The coefficient specification is highly consequential, since the signs of the coefficients of the structural causal variables of the model depend on the sign of the coefficient assigned to the normalizing indicator. Breusch points out that their conclusions – that increases in the tax burden, the size of government and the extent of self-employment all increase the size of the shadow economy – are a direct result of their arbitrary choice of a negative one coefficient for their normalizing indicator variable, real GDP.[27] Dell'Anno and Schneider (2006) confirm these findings and proceed to justify their arbitrary choices of the

[27] Dell'Anno's earlier working paper (Dell'Anno, 2003, p. 24) acknowledges the "strong dependence of outcomes by the choice of the coefficient of scale" and that "the signs of the determinants of the hidden economy … are a function of the researcher's choice". Unfortunately, these key admissions no longer appear in the published version of the paper (Dell'Anno & Schneider, 2003).

signs of the normalizing coefficients by appealing to "reductio ad absurdum" (Dell'Anno & Schneider, 2006, p.5; Dell'Anno, 2007, p.262). They state:

> When the "sign" of the coefficient of scale is changed from positive to negative, all the structural parameters of the causes change from positive to negative and vice versa (keeping the same absolute values), e.g. if a positive value is assigned to λ_1 [the scale coefficient] the relationship between tax burden and SE it becomes than negative [sic]. In our view, these results completely diverge from well-known theories and empirical results that assert a "positive" link between the underground economy and these variables (Dell'Anno & Schneider, 2006, p.5).

All of Schneider's MIMIC model papers conclude that higher tax rates increase the size of the shadow economy. However, tax evasion theory predicts that this relationship is either ambiguous (Allingham & Sandmo, 1972) or negative (Yitzhaki, 1974). Similarly, Schneider's arbitrary parametric choices force the conclusion that increased regulation unambiguously increases non-compliance. However, audit studies find that stricter income reporting regulations invariably improve compliance. Therefore, the results obtained by employing the conventional unit value as the normalizing coefficient are neither absurd, nor inconsequential. Schneider achieves consistent substantive results conforming to his prior beliefs by selecting indicator variables and normalization coefficients that vary from study to study. Table 3 lists the various indicator variables chosen for normalization in different papers, and their chosen coefficient.

Breusch (2005b, p.18) initially called attention to the fact that "the assignment of a negative coefficient to the normalizing indicator variable will reverse the sign of the latent variable. Since the latent variable is interpreted as a series of changes, that decision will *invert* the time path of the final result".[28]

One of the most bizarre, albeit largely unnoticed, changes to empirical results occurred when Schneider reversed his conclusions regarding the trend of the worldwide shadow economy. Schneider, Buehn, and Montenegro (2010a) initially reported estimates of the size of SSE as a percentage of GDP for 162 countries, concluding that, between 1999 and 2007, virtually all of the world's shadow economies *increased* in size.

[28] Dell'Anno's (2003, p.24) working paper hints at this inversion by stating that "if the parameter of scale is chosen to equal +1 (instead of -1) the estimated shadow economy became specular to time series displayed". However, Dell'Anno and Schneider's published version omits this observation, giving the impression that the negative relationship between the growth rate of GDP and the hidden economy is a result of the empirical findings rather than their arbitrary choice of a -1 normalizing coefficient (Dell'Anno & Schneider, 2003, pp.106, 112).

Table 3 MIMIC Indicator Variables and Normalization Coefficients

Paper	Page	Indicator Variable for Normalization	Coefficient
Dell'Anno & Schneider (2003)	102	Real GDP	-1
Bajada & Schneider(2005)	394	Currency holdings	1
Schneider (2005)	604	Annual rate of GDP change	-1
Schneider (2005)	605	Average working time (per week)	-1
Schneider (2007)	11	Annual rate of GDP	-1
Schneider (2007)	12	Average working time	-1
Buehn & Schneider (2008)	15	GDP	1
Buehn & Schneider (2008)	15	GDP	-1
Herwartz, Schneider, & Tafenau (2015)	1580	GDP per capita	-1
Schneider & Buehn (2009)	8	Growth rate of GDP	-1
Schneider & Buehn (2009)	10	GDP per capita	-1
Schneider & Buehn (2009)	11	GDP per capita	-1
Tafenau, Herwartz, & Schneider (2010)	632	GPD per capita	-1
Feld & Schneider (2010)	130	Average working time (per week)	-1
Schneider (2010)	450	Annual rate of GDP	1
Schneider. Buehn, & Montenegro (2010)	449	Currency	1

Note: The various models also include different additional indicator variables, some of which occasionally appear as causal variables, thereby violating the MIMIC specification requirement that the indicators are conditionally independent of the causes, given the latent variable.

Shortly thereafter, they produced a revised version of the paper, with the same title, models and parameter estimates (Schneider, Buehn, & Montenegro, 2010b), claiming that virtually all of the world's shadow economies had *decreased* in size during the same period.[29] This remarkable inversion of the time path results is mysteriously attributed to "a serious calibration error (sign switch)."[30] No further explanation is offered. The only clue as to what may have occurred appears as an inconspicuous addition of the following words to footnote 24 in the revised version of the paper: "The MIMIC index has been adjusted to the positive range by adding a positive constant."[31] My correspondence with the authors of the paper failed to provide any further clarification.

Subsequent versions of the Schneider, Buehn, and Montenegro results only confuse the matter further. Schneider and Enste (2013, Ch.4, p.37, Table 4.2) report the size and trend of the shadow economy in the 151 countries taken from the *original*

[29] All estimated parameters for each of the seven MIMIC model specifications reported in Version 1 (Schneider, Buehn, & Montenegro, 2010a) are identical to those of Version 2 (Schneider, Buehn, & Montenegro, 2010b); only the labels of the model specifications have changed. Nevertheless, the reported size of the shadow economy and its trend has changed for every one of the 162 countries.

[30] The lead footnote of the "revised" version (Schneider, Buehn, & Montenegro, 2010b) reads, "Unfortunately the estimates of the original version (WPS 5356) needed to be revised due to a serious calibration error (sign switch). We apologize for this, especially as we now have in this version a negative trend for the size and development of the shadow economies over 1999-2007, which we did not have in the original version."

[31] Schneider, Buehn, and Montenegro (2010c) make no mention of the "calibration error" but, in footnote 8, p.453, include reference to the adjustment required to make the MIMIC index positive.

version of Schneider, Buehn, and Montenegro (2010a, Table 3.3.6) as mostly increasing between 1999 and 2007. The same chapter (Schneider and Enste, 2013, Ch.4, p. 43, Table 4.3) includes results of the size and trend for 88 countries taken from the *revised* (Schneider, Buehn, & Montenegro, 2010b, Table 3.3.1) paper, showing a downward trend for the shadow economy.[32] Neither these nor other published versions of the same Schneider, Buehn, and Montenegro estimates (Buehn & Schneider, 2012a; Schneider & Williams, 2013) make mention of the "calibration error". These subsequent versions also omit mention of the mysterious addition of the positive constant required to make the MIMIC index positive.[33]

Another major point of contention regarding Schneider's MIMIC model estimates is their size. It is important to recognize that the MIMIC model produces only an index of SSE. Its size is determined by a separate "benchmarking" calibration procedure. Following Schneider, Buehn, and Montenegro (2010a, 2010b), the size of SSE (N^*_t) at time *t* is given as:

$$N^*_t = \tilde{N}_t/\tilde{N}_{2000} \times N^*_{2000}$$

where $\tilde{N}_t$ denotes the value of the MIMIC index at time t, $\tilde{N}_{2000}$ is the value of the index in the base year 2000, and N^*_{2000} is an exogenous estimate of SSE in the year 2000. The "exogenous" estimates for each of the 162 countries come from unspecified currency demand models. No documentation is given concerning the source or specification of each country's currency demand model required to produce the necessary N^*_{2000} "benchmark". This makes it impossible to determine either what each currency model was designed to measure, or the interval of uncertainty of the estimate.[34] Slemrod and Weber's (2012) critique of Schneider's benchmarking approach concludes that it "makes the estimates nearly impossible to interpret, since the estimates for each country are a function of other estimates, where the exact model used (by other researchers) to obtain these estimates are often directly violated in MIMIC" (Slemrod & Weber, 2012, p.49).

Breusch concludes his trenchant critique with a stern warning to the profession:

> The literature applying this model to the underground economy abounds with alarming Procrustean tendencies. Various kinds of sliding and scaling of the results are carried out in the name of "benchmarking", although these operations are not always clearly documented. The data are typically transformed in ways that are not only undeclared but have the unfortunate effect of making the results

[32] Both tables, taken from different versions of the paper, reference their source simply as Schneider, Buehn, and Montenegro (2010).

[33] Note its absence from footnote 17 in Buehn and Schneider (2012a, p.159).

[34] Breusch (2005b, p.28) and Slemrod and Weber (2012, pp.49-50) also demonstrate that, when currency is also used as an indicator along with a measure of income, the strict assumptions required of the correlation structure of the MIMIC model are violated. Ahumada, Alvaredo, and Canavese (2007) demonstrate how income elasticity estimates greater than or less than one in currency demand models will further bias these shadow economy estimates.

> of the study sensitive to the units in which the variables are measured. The complexity of the estimation procedure, together with its deficient documentation, leave the reader unaware of how these results have been shortened to fit the bed of prior belief. There are many other results in circulation for various countries, for which the data cannot be identified and which are given no more documentation than "own calculations by the MIMIC method". Readers are advised to adjust their valuation of these estimates accordingly (Breusch, 2005b, pp.28-29).

Despite these explicit and powerful warnings concerning the veracity of Schneider's MIMIC model results, his estimates have been repeatedly and inappropriately used as "the dependent variable in regression analyses in order to determine what causes noncompliance" (Slemrod & Weber, 2012, p.49). SSE estimates have now been regressed on virtually every conceivable variable for which comparable temporal cross-country data exist. These include: *tax morale* (Torgler & Schneider, 2009); *direct democracy* (Teobaldelli & Schneider, 2013); *unemployment* (Bajada & Schneider, 2009); *regulation* (Enste, 2010a, 2010b); *the quality of institutions* (Torgler & Schneider, 2009; Dreher, Kotsogiannis, & McCorriston, 2009; Dreher, Méon, & Schneider, 2014); *the influence of public institutions* (Schneider, 2010a); *corruption* (Schneider, 2007; Schneider & Buehn, 2009; Bovi & Dell'Anno, 2010; Buehn & Schneider, 2012b); *the official economy* (Dell'Anno, 2008); *enforcement* (Buehn & Schneider, 2012a); *voice, accountability and corruption (*Torgler, Schneider, & Macintyre, 2011); *work in the shadows* (Schneider, 2014); *energy prices* (Suslov & Ageeva, 2009); *decentralization* (Buehn, Lessmann, & Markwardt, 2013; Dell'Anno & Teobaldelli, 2015); *trust* (D'Hernoncourt & Méon, 2012); *education* (Buehn & Farzanegan, 2013); *pollution* (Elgin & Oztunali, 2014); *intelligence* (Salahodjaev, 2015*); inequality* (Dell'Anno, 2015); *religion* (Schneider, Linsbauer, & Heinemann, 2015); *internet usage* (Elgin, 2013); *quality of life* (Kireenko & Nevzorova, 2015) and *electronic payments* (Schneider, 2010b, 2013).

Commenting on such studies, Slemrod and Weber (2012, p.50) note:

> While estimates obtained from such an analysis may appear reasonable ex-post, they are not interpretable as estimates of any causal effect. They are useful neither for confirming ex-ante hypotheses nor for learning additional information about what factors cause the size of the informal economy to differ across countries (Slemrod & Weber, 2012, p.50).

A decade has passed since Breusch's critical evaluations of Schneider's earlier MIMIC model estimates. None of Schneider's subsequent studies using this

methodology has been, or can be, tested for robustness because, to date, he has not provided sufficient documentation for replication. What we have learned of Professor Schneider's work can be summarized as follows:

1) The ambiguous meaning of the latent variable estimated by the MIMIC model suggests that Schneider has not measured the entity he has defined; he has simply defined the entity he has measured.
2) His substantive conclusions concerning the effects of the causal variables on the size of the shadow economy are not determined by the data, but rather by his arbitrary choices of indicators and normalizing coefficients.
3) The downward trend he now reports for the world's shadow economies, between 1999 and 2007, results from an arbitrary and unexplained addition of a constant to the MIMIC index he originally calculated.
4) The magnitude of his shadow economy estimates are the result of his benchmarking the MIMIC index against currency demand model estimates of undocumented provenance, specifications of which typically violate the assumptions of the MIMIC model.
5) The proliferation of published studies regressing his unreliable shadow economy MIMIC estimates on virtually every other available temporal cross-country variable cast no light on the causes of the shadow economy, nor are they useful for testing any other ex-ante hypotheses.

The MIMIC model's complexity, the arbitrary procedures employed in its applications, the absence of information concerning the range of uncertainty associated with its estimates, and the lack of appropriate documentation required for replication lead one to the inescapable conclusion that Schneider's reported results are not credible estimates of any unobserved economy.

CONCLUSIONS

Every effort to observe and measure non-compliant behaviours confronts the social scientist with the analogue of the "observer effect" in physics, namely that the very act of observation affects the phenomenon being observed. The presence of any observer, be it a tax authority, a government agency charged with enforcing regulations, a statistical agency, or an experimental researcher (Milgram, 1963), leads individuals, households and firms to change their behaviour. Agents' attempts to avoid detection distort observation and make measurement difficult and costly. Nevertheless, measurement is essential if we are to understand the causes and the efficiency, equity and stabilization consequences of non-compliant behaviours. The distinguishing feature of each non-compliant behaviour is determined by the particular rule it violates, and its social, economic and political impact depends on the importance of the rule violated and the extent of the violation.

Early, crude attempts to estimate non-compliant economic activity suggested the existence of a growing unobserved economy sufficiently large to be of concern to both tax authorities and the custodians of the nation's information systems. In response, fiscal agencies set themselves the task of measuring the amount of revenue lost due to unreported income. Their studies confirm that substantial amounts of revenue are not collectable, but that the income categories subject to the strongest reporting regulations have the highest rates of compliance (IRS, 1973, 1983, 2005a, 2005b, 2005c, 2006, 2012). The theory of tax evasion predicts that improving compliance requires penalizing evaders and increasing probabilities of detection, while also refining the design of optimal tax systems to reduce the costs of compliance. Rule simplification, improved information and service provision, and enhancements in the efficiency and equity of public goods delivery systems all serve to enhance voluntary compliance (Slemrod, 1990; Alm, 1996; IRS, 2011; Sandmo, 2012). Recent efforts to assess the extent of revenue losses due to tax evasion involve both "bottom-up" tax gap estimates based on audits, surveys and data matching, and "top-down" estimates, the reliability of which depends critically on the exhaustiveness of the NIPA aggregates required for their construction (HM Revenue and Customs, 2014a, 2014b; European Commission, 2015).

In response to academic challenges, national and international agencies responsible for the collection and dissemination of macro-economic information now employ a coherent nomenclature describing the components of the NOE they seek to measure, as well as prescribed best practice methods for obtaining exhaustive measures of national income and product. Misreporting adjustments, accounting for unreported income missing from the tax return data used to construct components of the accounts, are included in measures of NOE. The statistical agencies of many countries now produce measures of the three major components of NOE, namely the underground, illegal and informal sectors. The complex inferential methods employed to estimate these non-observed components include modelling, surveys, and reconciliation of the supplies and uses of commodity flows.

Decisions concerning the allocation of private and public resources rely increasingly on the exhaustiveness, compatibility and accuracy of national information systems. In order to improve confidence in these information systems, as well as the outcomes of decisions based on them, recorded national accounts must be published on a timely basis, along with detailed estimates of both the observed and measured non-observed components of the accounts. Where possible, reporting should include estimates of error ranges to reflect associated uncertainty. The methods used to construct measures of the non-observed sector need to be transparent and strictly monitored for reliability and compatibility across countries and over time. To date, these confidence-building requirements have yet to be met for many countries, as we still lack consistent reports that document the manner and extent to which measures of NOE affect published national estimates of key

macroeconomic aggregates. The United Nations Economic Commission for Europe could rectify this situation by collecting and systematically publishing this information in an expanded and updated edition of its *Survey of Country Practices* (see UN, 2003, 2008).

MIMIC model applications treating the "shadow economy" as a latent variable purport to measure the underground component of NOE. Analyses of these applications reveal that the statistical and economic assumptions of the MIMIC model are typically violated, and that the resulting latent variable has little relationship with any unobserved economy. The methodology has been shown to be so malleable that it can be readily manipulated to obtain virtually any desired result; however, the complexity of the procedure often obscures these manipulations.

The major proponent of this arcane methodology is Professor Friedrich Schneider, who claims to have estimated the size and trend of the shadow economy worldwide (Schneider, 2005, 2007; Schneider & Buehn, 2009; Schneider, Buehn, & Montenegro, 2010a, 2010b, 2010c, 2011). If the veracity of his results were substantiated, they would represent an important contribution to the field. To date, insufficient and inaccurate documentation concerning key data sources and procedures has precluded replication of his results. However, we know that his estimates critically depend on his choice of indicator variables and the sign of their normalizing coefficients, on which theory provides little guidance. The size and trend of his latent variable is arbitrary and fragile, its meaning is obscure, and his estimates bear no relation to existing national accounting measurements of the NOE component he claims to estimate. It is time to acknowledge that both the conceptual and empirical basis of Schneider's shadow economy are insubstantial. The repeated use of his flawed MIMIC estimates as dependent variables in subsequent studies is empirically unjustified. These fundamental defects of the MIMIC applications documented by various researchers are increasingly acknowledged and cited by Schneider, but he chooses to ignore the implication of these critiques,[35] namely that the estimates he continues to present are untenable and do not contribute to our stock of useful knowledge.

It does not speak well for our profession that these conceptually faulty, highly manipulated, and largely undocumented estimates continue to be published in our academic journals. This practice would be curtailed if all journal editors were to adopt and enforce the submission guidelines and data availability policies required by the American Economics Association. It is also unfortunate that our literature continues to confound various unobserved economies, without carefully

[35] Responding to Breusch's (2005b) critique, Dell'Anno and Schneider (2006, p.17) conclude that "the MIMIC model is still one of the best approaches to this purpose".

distinguishing which set of institutional rules are being violated by the behaviour of concern. Tax evasion, the NOE, the illegal economy, the corruption economy and the illegal alien economy may overlap to some extent; however, their nature, measurement, consequences and policy implications are quite different. Policy makers are cautioned neither to blur these distinctions nor to be influenced by unsubstantiated estimates of the so-called "shadow economy". The popular press must become more circumspect about citing them uncritically. It is time to bring greater credibility to the study of unobserved economies by acknowledging that, to date; distinctions between them have too often been ignored, and that the results obtained by prevailing applications of the MIMIC methodology are unworthy of a place in the academic, policy and popular literature. Further research must begin with a greater willingness to acknowledge the critical limitations of what we too often claim to know.

Future research must continue the quest to understand the nature, causes, consequences and extent of non-compliant behaviors. Extensions of the theory of tax evasion demonstrate how individual decisions regarding tax compliance may affect the performance of the entire economic system (Sandmo, 2012). Reliable empirical estimates of the extent, trend and costs of non-compliance are required to mobilize the public resources necessary to deal with its consequences. Greater creativity and inventiveness is necessary to develop unobtrusive measures of non-compliant behaviours and a deeper understanding of the traces that these behaviours leave behind.[36] Cash and, more recently, virtual currencies, being preferred media of exchange for suspect transactions, provide promising clues to trends in non-compliant activities.[37]

If MIMIC applications are to be employed, their cause/indicator structure must be consistent with both the statistical assumptions of the latent variable specification and with relevant economic theory pertaining to the particular non-compliant behaviour being estimated. Calibration methods must only employ exogenous estimates, such as point estimates of tax gaps independently derived from audit studies, or exogenous national accounts estimates of measured NOE. Above all, macro approaches to measurement must adhere to stricter standards of transparency through full provision of data, sources, transformations, statistical assumptions, estimation methods, pre-testing biases and error margins of results. All publishable studies must be readily replicable in order to test the reliability and robustness of the findings.

[36] For example, Pissarides and Weber (1989) and Feldman and Slemrod (2007) rely on anomalies in food expenditure and charitable contribution patterns, respectively, as traces of misreported income.
[37] Unexplained changes in per-capita cash holdings and changes in the velocity of cash, as evidenced by changes in the average lifetimes of note denominations, yield potential traces of non-compliant activities (Feige, 1989b).

Ultimately, a preeminent goal of social policy is the achievement of greater voluntary compliance with good rules. Creative approaches are needed to reduce compliance and administrative costs, target appropriate deterrence measures, improve the perceived equity of institutional rules, enhance the efficiency and quality of public goods delivery, and innovatively restructure choice architecture through increased reliance on "nudges" (Thaler & Sunstein, 2008). Non-compliance research is not limited to violations of fiscal rules and conventions of national income accounting. The profession faces major challenges in its efforts to observe, measure and understand the causes and consequences of non-compliant behaviours involving undocumented workers, illegal immigrants, human, drugs and arms traffickers, and planetary polluters. These issues will continue to test the theoretical, observational and measurement skills of the social science community for many years to come.

REFERENCES

Ahumada, H., Alvaredo, F., & Canavese, A. (2007). The monetary method and the size of the shadow economy: A critical assessment. *Review of Income and Wealth, 53*(2), 363-371.

Allingham, M.G., & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics, 1*, 323-338.

Alm, J. (1996). What is an "optimal" tax system? *National Tax Journal, 49*(1), 117-133.

Bajada, C., & Schneider, F. (2005). The shadow economies of the Asia-Pacific. *Pacific Economic Review, 10*(3), 379-401.

Bajada, C., & Schneider, F. (2009). Unemployment and the shadow economy in the OECD. *Revue Économique, 60*(5), 1033-1067.

BEA (2009). *Concepts and Methods of the US National Income and Product Accounts*, Chapters 1-5. Washington, DC: Bureau of Economic Analysis, US Department of Commerce.

Black, T., Bloomquist, K., Emblom, E., Johns, A., Plumley, A., & Stuk, E. (2012). *Federal Tax Compliance Research: Tax Year 2006 Tax Gap Estimation*. Washington, DC: IRS Research, Analysis & Statistics.

Blinder, A. S., & Rudd, J. B. (2012). The supply-shock explanation of the great stagflation revisited. In M. D. Bordo, & A. Orphanides (Eds), *The Great Inflation: The Rebirth of Modern Central Banking* (pp.119-180). Chicago, IL: University of Chicago Press.

Bollen, K. A. (1989). *Structural Equations with Latent Variables.* New York: John Wiley & Sons.

Bovi, M., & Dell'Anno, R. (2010). The changing nature of the OECD shadow economy. *Journal of Evolutionary Economics, 20*(1), 19-48.

Breusch, T. (2005a). The Canadian underground economy: An examination of Giles and Tedds. *Canadian Tax Journal, 53*(2), 367-391.

Breusch, T. (2005b). *Estimating the Underground Economy using MIMIC Models*. Australian National University Working Paper, Canberra. Retrieved from http://econ.unimelb.edu.au/SITE/workshops/Breusch.pdf.

Breusch, T. (2005c). *Fragility of Tanzi's Method of Estimating the Underground Economy*. Working Paper, Center for Social Research, The Australian National University, Canberra.

Breusch, T. (2006a). Book review of "Size, Causes and Consequences of the Underground Economy: An International Perspective". *The Economic Record, 82*(259), 492-494.

Breusch, T. (2006b) *Australia's Underground Economy: Redux?* Australian National University Working Paper, Canberra. Retrieved from http://mpra.ub.uni-muenchen.de/9980/.

Buehn, A., & Farzanegan, M. R. (2013). Impact of education on the shadow economy. *Economics Bulletin, 33*(3), 2053-2063.

Buehn, A., Lessmann, C., & Markwardt, G. (2013). Decentralization and the shadow economy: Oates meets Allingham-Sandmo. *Applied Economics, 45*(18), 2567-2578.

Buehn, A., & Schneider, F. (2008). *MIMIC Models, Cointegration and Error Correction: An Application to the French Shadow Economy*. CESifo Working Paper No. 2200, CESIfo, Munich.

Buehn, A., & Schneider, F. (2012a). Shadow economies around the world: Novel insights, accepted knowledge, and new estimates. *International Tax and Public Finance, 19,* 139-171.

Buehn, A., & and Schneider, F. (2012b). Corruption and the shadow economy: Like oil and vinegar, like water and fire? *International Tax and Public Finance, 19,* 172-194.

Cagan, P. (1958). The demand for currency relative to the total supply of money. *Journal of Political Economy, 66*(3), 302-328.

Campbell, D. T. (1975). Assessing the impact of planned social change. In G. Lyons (Ed.), *Social Research and Public Policies: The Dartmouth/OECD Conference* (pp.3-45). Hanover, NH: The Public Affairs Center, Dartmouth College.

Carpenter, J., & Matthews, P. (2005). *Norm Enforcement: Anger, Indignation or Reciprocity?* Middlebury College Economics Discussion Paper No. 05-03, Middlebury College, VT.

Carson, C. (1984). The underground economy: An introduction. *Survey of Current Business, 64,* 21-37.

Charmes, J. (2012). The informal economy worldwide: Trends and characteristics. *The Journal of Applied Economic Research, 6*(2), 103-132.

Charmes, J. (2014). *How to Incorporate Informal Sector Data into National Accounts: A Guidebook.* Addis Ababa: Africa Centre of Statistics, UNECA.

Ciupagea, C. (2004). Contribution of the informal activities to the Romanian economy. In R. Neef, & P. Adair (Eds), *Informal Economies and Social Transformation in Romania* (pp.145-154). Münster: LIT Verlag.

Cowell, F. (1990). *Cheating the Government.* Cambridge, MA: MIT Press.

Dell'Anno, R. (2003). *Estimating the Shadow Economy in Italy: A Structural Equation Approach*. Working Paper No. 2003-07, University of Aarhus, Denmark.

Dell'Anno, R. (2007). The shadow economy in Portugal: An analysis with the MIMIC approach. *Journal of Applied Economics, 10*(2), 253-277.

Dell'Anno, R. (2008). What is the relationship between unofficial and official economy? An analysis in Latin American countries. *European Journal of Economics, Finance and Administrative Sciences, 12,* 185-203.

Dell'Anno, R. (2015). Analyzing the determinants of the shadow economy with a "separate approach": An application of the relationship between inequality and the shadow economy. *World Development* [online]. doi:10.1016/j.worlddev.2015.08.026.

Dell'Anno, R., & Schneider, F. (2003). The shadow economy of Italy and other OECD countries: What do we know? *Journal of Public Finance and Public Choice, 21*(2-3), 97-120.

Dell'Anno, R., & Schneider, F. (2006). *Estimating the Underground Economy using MIMIC Models: A Response to T. Breusch's Critique*. Working Paper No. 0607, Department of Economics, Johannes Kepler University of Linz, Linz, Austria.

Dell'Anno, R., & Teobaldelli, D. (2015). Keeping both corruption and the shadow economy in check: The role of decentralization. *International Tax and Public Finance, 22*(1), 1-40.

D'Hernoncourt, J., & Méon, P. J. (2012). The not so dark side of trust: Does trust increase the size of the shadow economy? *Journal of Economic Behavior & Organization, 81*, 97-121.

Dobozi, I., & Pohl, G. (1995). Real output decline in transition economies: Forget GDP, try power consumption data. *Transition Newsletter, 6*(1-2), 17-18.

Dreher, A., Kotsogiannis C., & McCorriston, S. (2009). How do institutions affect corruption and the shadow economy? *International Tax and Public Finance, 16*(6), 773-796.

Dreher, A., Méon, P. J., & Schneider, F. (2014). The devil is in the shadow: Do institutions affect income and productivity or only official income and official productivity? *Public Choice, 158*(1), 121-141.

Eilat, Y., & Zinnes, C. (2002). *The Evolution of the Shadow Economy in Transition Countries: The Consequences for Economic Growth and Donor Assistance*. CAER II Discussion Paper No. 8, Harvard Institute for International Development, Cambridge, MA.

Elgin, C. (2013). Internet usage and the shadow economy: Evidence from panel data. *Economic Systems, 37*(1), 111-121.

Elgin, C., & Oztunali, O. (2014). Pollution and informal economy. *Economic Systems, 38*(3), 333-349.

Enste, D. (2010a). Regulation and the shadow economy. *Constitutional Political Economy, 21*(3), 231-248.

Enste, D. (2010b). Shadow economy: The impact of regulation in OECD-countries. *International Economic Journal, 24*(4), 555-571.

Epstein, R. (1995). *Simple Rules for a Complex World.* Cambridge, MA: Harvard University Press.

Ericson, R. (1984). The "second economy" and resource allocation under central planning. *Journal of Comparative Economics, 8*(1), 1-24.

Center for Social and Economic Research (2013a). *2012 Update Report to the Study to Quantify and Analyze the VAT Gap in the EU-27 Member States*. European Commission TAXUD/2013/DE/321, CPB Netherlands Bureau for Economic Policy Analysis, The Hague, Netherlands.

Center for Social and Economic Research (2013b). *Study to Quantify and Analyze the VAT Gap in the EU-27 Member States*. European Commission

TAXUD/2012/DE316, CPB Netherlands Bureau for Economic Policy Analysis, The Hague, Netherlands.

Economist (2014, 31 May). Sex, drugs and GDP. *The Economist*. Retrieved from http://www.economist.com/news/finance-and-economics/21603073-italys-inclusion-illicit-activities-its-figures-excites-much-interest-sex.

ESA (1996). Council Regulation (EC) No 2223/93 of 25 June 1996 on the European system of national and regional accounts in the Community. *Official Journal of the European Communities*, L310/1.

European Commission (2015). *Study to Quantify and Analyze the VAT Gap in the EU Member States: 2015 Report*. European Commission TAXUD/2013/DE321, CPB Netherlands Bureau for Economic Policy Analysis, The Hague, Netherlands.

Feige, E. L. (1979). How big is the irregular economy? *Challenge, 22*(1), 5-13.

Feige, E. L. (1980). Macroeconomics and the unobserved sector. In W. Block, & M. Walker (Eds), *Taxation: An International Perspective* (pp.21-26). Vancouver, Canada: The Fraser Institute.

Feige, E. L. (1986). A reexamination of the "underground economy" in the United States: A comment on Tanzi. *International Monetary Fund Staff Papers, 33*(4), 768-781.

Feige, E. L. (1989a). *The Underground Economies: Tax Evasion and Information Distortion.* Cambridge: Cambridge University Press.

Feige, E. L. (1989b). *Currency Velocity and Cash Payments in the US Economy: The Currency Enigma*. Munich Personal RePEc Archive Paper 13807. Retrieved from http://mpra.ub.uni-muenchen.de/13807/.

Feige, E. L. (1990). Defining and estimating underground and informal economies: The New Institutional Economics approach. *World Development, 18*(7), 989-1002.

Feige, E. L. (2003a). Dynamics of currency substitution, asset substitution and de facto dollarization and euroization in transition countries. *Comparative Economic Studies, 45*(3), 358-383.

Feige, E. L. (2003b). Unofficial dollarization in Latin America: Currency substitution, network externalities and irreversibility. In J. Dean, D. Salvatore, & T. Willett (Eds), *The Dollarization Debate* (pp.36-71). Oxford: Oxford University Press.

Feige, E. L. (2012a). New estimates of US currency abroad, the domestic money supply and the unreported economy. *Crime, Law and Social Change, 57*(3), 239-263.

Feige, E. L. (2012b). The myth of the "cashless society": How much of America's currency is overseas? In Deutsche Bundesbank Eurosystem, *International Cash Conference 2012: The Usage, Costs and Benefits of Cash: Theory and Evidence from Macro and Micro Data* (pp.119-151). Frankfurt am Main: Deutsche Bundesbank.

Feige, E. L., & Urban, I. (2003). *Estimating the Size and Growth of Unrecorded Economic Activity in Transition Countries: A Re-evaluation of Electric*

Consumption Method Estimates and their Implications. William Davidson Institute Working Paper Number 636. Retrieved from http://www.bus.umich.edu/KresgeLibrary/Collections/Workingpapers/wdi/wp636.pdf.

Feige, E. L., & Urban, I. (2008). Measuring underground (unobserved, non-observed, unrecorded) economies in transition countries: Can we trust GDP? *Journal of Comparative Economics, 36*(2), 287-306.

Feld, L., & Schneider, F. (2010). Survey on the shadow economy and undeclared earnings in the OECD countries. *German Economic Review, 11*(2), 109-149.

Feldbrugge, F. J. (1989). The Soviet second economy in a political and legal perspective. In E. L. Feige (Ed.), *The Underground Economies: Tax Evasion and Information Distortion* (pp.297-338). New York: Cambridge University Press.

Feldman, N.E., & Slemrod, J. (2007). Estimating tax noncompliance with evidence from unaudited tax returns. *The Economic Journal*, 117, 327-352.

Frey, B. S., & Weck-Hannemann, H. (1984). The hidden economy as an "unobserved" variable. *European Economic Review, 26*(1-2), 33-53.

Friedman, E., Johnson, S., Kaufmann, D., & Zoido-Lobaton P. (2000). Dodging the grabbing hand: The determinants of unofficial activity in 69 countries. *Journal of Public Economics, 76*(3), 459-493.

Gemmell, N., & Hasseldine J. (2012). The tax gap: A methodological review. *Advances in Taxation, 20,* 203-231.

Giles, D. (1999). Modeling the hidden economy and the tax gap in New Zealand. *Empirical Economics, 24*, 621-640.

Giles, D., & Tedds, L. (2002). *Taxes and the Canadian Underground Economy*, Toronto: Canadian Tax Foundation.

Golovko, V. (n.d.). *Experience of Ukraine in Measuring of Non-Observed Economy.* State Statistics Committee, Ukraine.

Grossman, G. (1977). The second economy in the USSR. *Problems of Communism, 26*(5), 25-40.

Gutmann, P. (1977). The subterranean economy. *Financial Analyst Journal, 34*(1), 24-27.

Gyomai, G., & Van de Ven, P. (2014). The non-observed economy in the system of national accounts. *OECD Statistics Brief, 18*, 1-12

Helberger, C., & Knepel, H. (1988). How big is the shadow economy? A re-analysis of the unobserved-variable approach of B. S. Frey and H. Weck-Hannemann. *European Economic Review, 32*, 965-976.

Herwartz, H., Schneider F., & Tafenau E. (2015). One share fits all? Regional variations in the extent of the shadow economy in Europe. *Regional Studie*s, *49*, 1575-1587.

Hill, R. (2002). The underground economy in Canada: Boom or bust? *Canadian Tax Journal, 50*(5), 1641-1654.

HM Revenue and Customs (2014a). *Measuring Tax Gaps, 2014 Edition: Tax Gap Estimates for 2012-13*. London: HMSO.

HM Revenue and Customs (2014b). *Measuring Tax gaps, 2014 Edition: Methodological Annex*. London: HMSO.

IRS (1979). *Estimates of Income Unreported on Individual Tax Returns*. Washington, DC: US Department of the Treasury.

IRS (1983). *Income Tax Compliance Research 1983: Estimates for 1973-1981*. Washington, DC: US Department of the Treasury.

IRS **(**2005a). New IRS study provides preliminary tax gap measure. *IRS Newswire*, IR-2005-38. Washington, DC: US Department of the Treasury.

IRS **(**2005b). *Tax Gap Facts and Figures*. Washington, DC: US Department of the Treasury.

IRS **(**2005c). Understanding the tax gap. *IRS Newswire*, FS-2005-14. Washington, DC: US Department of the Treasury.

IRS **(**2006). *Updated Estimates of the TY 2001 Individual Income Tax Underreporting Gap: Overview*. Washington, DC: Office of Research, Analysis, and Statistics, US Department of the Treasury.

IRS (2011). Recent research on tax administration and compliance. *IRS Research Bulletin*, 1500. Retrieved from https://www.irs.gov/pub/irs-soi/11rescon.pdf.

IRS (2012). *Tax Year 2006 Tax Gap Estimate: Summary of Estimation Methods*. Washington, DC: US Department of the Treasury. Retrieved from http://www.irs.gov/pub/irs-soi/06rastg12methods.pdf.

IMF (2013). *United Kingdom Assessment of HMRC's Tax Gap Analysis*. IMF Country Report No. 13/314, Fiscal Affairs Department, International Monetary Fund.

IMF (2014). *Republic of Estonia Technical Assistance Report-Revenue Administration Gap Analysis Program: The Value Added Tax Gap*. IMF Country Report No.14/13, Fiscal Affairs Department, International Monetary Fund.

ISWGNA (1993). *System of National Accounts 1993*. New York: Inter-Secretariat Working Group on National Accounts. Retrieved from http://unstats.un.org/unsd/nationalaccount/ docs/1993sna.pdf.

Johnson, S., Kaufmann, D., & Schleifer, A. (1997). The unofficial economy in transition. *Brookings Papers on Economic Activity*, *2*, 159-239.

Johnson, S., Kaufmann, D., & Zoido-Lobaton, P. (1998). Regulatory discretion and the unofficial economy. *American Economic Review, 88*(2), 387-392.

Kaufmann, D., & Kaliberda, A. (1996). *Integrating the Unofficial Economy into the Dynamics of Post-Socialist Economies*. World Bank Policy Research Working Paper No. 1691, World Bank, Washington, DC.

Khomenko, T. (2007). *Measurement of the Non-Observed Economy in Transition Economies with Special Reference to the CIS countries*. Paper presented at the International Conference organized by the International Association for

Research in Income and Wealth (IARIW) and the National Bureau of Statistics (NBS) of China, Beijing, China, 18-21 September.

Kireenko, A., & Nevzorova, E. (2015). Impact of shadow economy on quality of life: Indicators and model selection. *Procedia Economics and Finance*, *25*, 559-568.

Kudabaev, Z. (2004). *Estimation of Non-Observed Economy in Kyrgyzstan.* Paper presented at the OECD/ESCAP/ADB Workshop on Assessing and Improving Statistical Quality: Measuring the Non-Observed Economy, Bangkok, 11-14 May.

Ledbetter, M. (2007). Comparison of BEA estimates of personal income and IRS estimates of adjusted gross income: New estimates for 2005, revised estimates for 2004. *Survey of Current Business*, November, 35-40.

Leamer, E. (1983). Let's take the con out of econometrics. *The American Economic Review*, *73*(1), 31-43.

Leitzel, J. (1997). Rule evasion in transitional Russia. In J. M. Nelson, C. Tilly, & L. Walker (Eds), *Transforming Post-Communist Political Economies* (pp.118-130). Washington, DC: National Academy Press.

Loayza, N. A. (1997). *The Economics of the Informal Sector: A Simple Model and Empirical Evidence from Latin America*. World Bank Policy Research Working Paper 1727, World Bank, Washington, DC.

Manski, C. F. (2015). Communication uncertainty in official economic statistics: An appraisal fifty years after Morgenstern. *Journal of Economic Literature, 53*(3), 631-653.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, *67*(4), 371-378.

Mosakova, I. (2001). *Estimation of Non-Observed Economy: The Statistical Practices in Russia*. Paris: OECD. Retrieved from http://www.oecd.org/std/na/2464219.pdf.

Murphy, R. (2012). *Closing the European Tax Gap: A Report for Group of the Progressive Alliance of Socialists & Democrats in the European Parliament*. Downham Market: Tax Research LLP. Retrieved from http://www.socialistsanddemocrats.eu/publications/closing-european-tax-gap-report-group-progressive-alliance-socialists-democrats.

OECD (2002). *Measuring the Non-Observed Economy: A Handbook*. Paris: OECD Publications.

ONS (2014). *National Accounts Articles: Impact of ESE95 Changes on Current Price GDP Estimates*. London: Office for National Statistics. Retrieved from www.ons.gov.uk/ons/dcp171766_365274.pdf.

Parker, R. (1984). Improved adjustments for misreporting of tax return information used to estimate the national income and product accounts. *Survey of Current Business*, June, 17-25.

Pissarides, C. A., & Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics, 39*(1), 17-32.

Putniņš, T. J., & Sauka, S. (2015). Measuring the shadow economy using company managers. *Journal of Comparative Economics*, *43*(2), 471-490.

Reckon LLP (2009). *Study to Quantify and Analyze the VAT Gap in the EU-25 Member States*. London: Reckon LLP.

Rogoznikova, T. (2004). *Experience of Uzbekistan in Measuring the Non-Observed Economy, NOE*. Paper presented at the Issues of Measurement Conference, St. Petersburg (in Russian).

Salahodjaev, R. (2015). Intelligence and shadow economy: A cross-country empirical assessment. *Intelligence*, *49*, 129-133.

Sandmo, A. (2012). An evasive topic: Theorizing about the hidden economy. *International Tax and Public Finance, 19*, 5-24.

Schneider, F. (2005). Shadow economies around the world: What do we really know? *European Journal of Political Economy*, *21*, 598-642.

Schneider, F. (2007). Shadow economies and corruption all over the world: New estimates for 145 countries. *Economics eJournal*, 2007-9.

Schneider, F. (2009). Size and development of the shadow economy in Germany, Austria and other OECD countries. *Revue Économique, 60*(5), 1079-1116.

Schneider, F. (2010a). The influence of public institutions on the shadow economy: An empirical investigation for OECD countries. *Review of Law and Economics*, *6*(3), 441-468.

Schneider, F. (2010b) *The Shadow Economy in Europe, 2010: Using Electronic Payments Systems to Combat the Shadow Economy*. Chicago, IL: A.T. Kearney. Retrieved from http://media.hotnews.ro/media_server1/document-2011-05-8-8602544-0-shadow.pdf.

Schneider, F. (2011). The shadow economy labour force: What do we (not) know? *World Economics*, *12*(4), 53-92.

Schneider, F. (2012). The shadow economy and tax evasion: What do we (not) know? Tax evasion in Europe (focus). *CESifo Forum*, *13*(2), 3-12.

Schneider, F. (2013). *The Shadow Economy in Europe, 2013*. Chicago, IL: A.T. Kearney. Retrieved from https://www.atkearney.com/documents/10192/1743816/The+Shadow+Economy+in+ Europe+2013.pdf.

Schneider, F. (2014). Work in the shadow: Micro and macro results. *International Economic Journal*, *28*(3), 365-379.

Schneider, F., & Bajada, C. (2005). An international comparison of underground economic activity. In C. Bajada, & F. Schneider (Eds), *Size, Causes and Consequences of the Underground Economy: An International Perspective* (pp.73-106). Aldershot: Ashgate Publishing Limited.

Schneider, F., & Buehn, A. (2009). Shadow economies and corruption all over the world: Revised estimates for 120 Countries. *Economics eJournal*, Version 2, 27 October.

Schneider, F., Buehn, A., & Montenegro, C. E. (2010a). *Shadow Economies All over the World: New Estimates for 162 Countries from 1999 to 2007*. World

Bank Policy Research Working Paper 5356, The World Bank, Washington, DC.

Schneider, F., Buehn, A., & Montenegro, C. E. (2010b). *Shadow Economies All over the World: New Estimates for 162 Countries from 1999 to 2007* (Revised Version). World Bank Policy Research Working Paper 5356, The World Bank, Washington, DC.

Schneider, F., Buehn, A., & Montenegro, C. E. (2010c). New estimates from the shadow economies all over the world. *International Economic Journal*, *24*(4), 443-461.

Schneider, F., Buehn, A., & Montenegro, C. E. (2011). Shadow economies all over the world: New estimates for 162 countries from 1999 to 2007. In F. Schneider (Ed.), *Handbook on the Shadow Economy*. Cheltenham: Edward Elgar.

Schneider, F., & Enste, D. (2000a). *Shadow Economies Around the World: Size Causes and Consequences*. IMF Working Paper WP/00/26, International Monetary Fund, Washington, DC.

Schneider, F., & Enste, D. (2000b). Shadow economies: Size, causes, and consequences. *Journal of Economic Literature, 38,* 77-114.

Schneider, F., & Enste, D. (2002). *Hiding in the Shadows: The Growth of the Underground Economy*. Economic Issues No. 30, International Monetary Fund, Washington, DC.

Schneider, F., & Enste, D. (2013). *The Shadow Economy: An International Survey*, second edition. New York: Cambridge University Press.

Schneider, F., Linsbauer, K., & Heinemann, F. (2015). Religion and the shadow economy. *Kyklos*, *68*(1), 111-141.

Schneider, F., & Neck, R. (1993). The development of the shadow economy under changing tax systems and structures: Some theoretical and empirical results for Austria. *Finanzarchiv*, *50*(3), 344-369.

Schneider, F., Raczkowski, K., & Mróz, B. (2015). Shadow economy and tax evasion in the EU. *Journal of Money Laundering Control*, *18*(1), 34-51.

Schneider, F., & Williams, C. (2013). *The Shadow Economy*. London: Institute of Economic Affairs.

Slemrod, J. (1990). Optimal taxation and optimal tax systems. *Journal of Economic Perspectives*, *4*(1), 157-178.

Slemrod, J. (2004). Small business and the tax system. In H. Aaron, & J. Slemrod (Eds), *The Crisis in Tax Administration* (pp.69-101). Washington, DC: Brookings Institution Press.

Slemrod, J. (2007). Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives*, *21*(1), 25-48.

Slemrod, J., & Weber, C. (2012). Evidence of the invisible: Toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance*, *19*(1), 25-53.

Smith, R. S. (2002). The underground economy: Guidance for policy makers? *Canadian Tax Journal, 50*(5), 1655-1661.

Strümpel, B. (1966). *Steuermoral und Steuerwiderstand der deutschen Selbständigen*. Cologne/Opladen: Westdeutscher Verlag.

Suslov, N., & Ageeva, S. (2009). *Influence of Energy Prices on the Size of Shadow Economy: A Cross-Country Analysis*. EERC Working Paper Series 09/05e, EERC Research Network, Kyiv, Ukraine.

Tafenau, E., Herwartz, H., & Schneider, F. (2010). Regional estimates of the shadow economy in Europe. *International Economic Journal*, *24*(4), 629-636.

Tanzi, V. (1983). The underground economy in the United States: Annual estimates, 1930-80. *International Monetary Fund Staff Papers*, 30. doi:10.2307/3867001.

Tedds, L. M. & Giles, D. E. A. (2005). Response to Breusch's critique. *Canadian Tax Journal*, *53*(2), 392-395.

Teobaldelli, D., & Schneider, F. (2013). The influence of direct democracy on the shadow economy. *Public Choice*, *157*, 543-567.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness.* New Haven, CT: Yale University Press.

Thomas, J. J. (1992). *Informal Economic Activity*. Ann Arbor, MI: The University of Michigan Press.

Torgler, B., & Schneider, F. (2009). The impact of tax morale and institutional quality on the shadow economy. *Journal of Economic Psychology, 30*, 228-245.

Torgler, B., Schneider, F., & Macintyre, A. (2011). Shadow economy, voice and accountability and corruption. In F. Schneider (Ed.), *Handbook on the Shadow Economy*. Cheltenham: Edward Elgar.

Torgler, B., Schneider, F., & Schaltegger, C. (2010). Local autonomy, tax morale and the shadow economy. *Public Choice, 144*(1), 293-321.

UN Economic Commission for Europe (2003). *Non-Observed Economy in National Accounts: Survey of National Practices*. Geneva: United Nations.

UN Economic Commission for Europe (2008). *Non-Observed Economy in National Accounts: Survey of Country Practices*. Geneva: United Nations.

Van de Ven, P. (2015). New standards for compiling national accounts: What's the impact on GDP and other macro-economic indicators? *OECD Statistics Brief,* 20. Retrieved from https://www.oecd.org/std/na/new-standards-for-compiling-national-accounts-SNA2008-OECDSB20.pdf.

Yitzhaki, S. (1974). A note on income tax evasion: A theoretical analysis. *Journal of Public Economics*, *3*, 201-202.

ISSN 2059-190X (Online)

Estimating the Underground Economy using MIMIC Models

Trevor Breusch[38]

Abstract

MIMIC models are being used to estimate the size of the underground economy or the tax gap in various countries. In this paper, I examine critically both the method in general, and three applications of the method by Giles and Tedds (2002), Bajada and Schneider (2005) and Dell'Anno and Schneider (2003). Connections are shown with familiar econometric models of linear regression and simultaneous equations. I also investigate the auxiliary procedures used in this literature, including differencing as a treatment for unit roots and the calibration of results using other data. The three applications demonstrate how the method is subjective and pliable in practice. I conclude that the MIMIC method is unfit for the purpose.

INTRODUCTION

By definition, the underground economy cannot be directly observed so its magnitudes must be estimated. Many different methods are employed for this purpose. Tax audits are informative, but are usually targeted toward suspected offenders, and hence are biased estimators of aggregate behaviour. Regular surveys of household expenditures and incomes conducted by national statistical agencies can be examined for discrepancies that might indicate unreported incomes. Special surveys are sometimes conducted, with direct questions about below-the-counter incomes or cash payments, although non-response bias is always a concern. At a more aggregate level, inferences can be made from inconsistencies between expenditure, income and product data collected from various sources for national accounting purposes. The most popular methods in the academic literature are based on macroeconomic models of either the demand for currency holdings (perhaps in comparison with bank account balances) or the consumption of some standard commodity such as electricity.

Interest is burgeoning in a more complex approach known as the "structural equation" or "multiple indicator multiple cause" (MIMIC) model. This method has its origins in the factor analysis literature of psychometrics, while its exposure in economics has been through the latent variable models of Zellner (1970) and Goldberger (1972). In the first application of MIMIC to estimating the underground economy, Frey and Weck-Hannemann (1984) examine a pooled dataset from 17 OECD countries. The idea is extended by Aigner, Schneider and Ghosh (1988), who allow some lagged adjustment in a dynamic MIMIC (or DYMIMIC) model and apply the method to the US. Giles (1999) further modifies the approach to incorporate developments in time-series methods, especially unit roots and cointegration analysis, and provides estimates of New Zealand's hidden economy.

[38] Crawford School of Public Policy, The Australian National University

The state of the art of dynamic MIMIC modelling is given by Giles and Tedds (2002), who describe the approach in detail and apply it to Canada. Authors taking up the method in the wake of Giles and Tedds include Bajada and Schneider (2005), who study Australia and other Pacific nations, and Dell'Anno and Schneider (2003), who estimate the underground economy in Italy and report results for other OECD countries.

The MIMIC approach is attractive in this context. The idea is to represent the output (or income) of the underground economy as a latent variable or index, which has causes and effects that are observable but which cannot itself be directly measured. Thus, there are two kinds of observed variable in the model, "causal" variables and "indicator" variables, which are connected by a single unobserved index. Values of the index over time are inferred from data on causes and indicators by estimating the statistical model and predicting the index. The fitted index is then interpreted as a time-series estimate of the magnitude of the underground economy. Usually, the measure is hidden output or income as a percentage of recorded GDP, although some researchers are concerned with the "tax gap" between actual and potential revenue when all taxable income is reported.

Bold claims are made by proponents of these methods for their ability to measure hidden economic activity. The estimates in the literature are often presented to three or four digits of precision and without any interval of uncertainty. The estimates are always large enough to cause grave concern and attract media headlines, and often the underground economy is shown to be growing strongly. These results have serious implications for economic and social policy in the areas of tax administration, national income accounting, stabilization policy, and social fairness and cohesion.

This use of MIMIC modelling has its critics. Helberger and Knepel (1988) show that the pioneering results of Frey and Weck-Hannemann are unstable in the face of minor changes in either the data period or the group of countries studied. They also argue that the lists of causal and indicator variables are unconvincing for the purpose. Smith (2002) and Hill (2002) criticise Giles and Tedds' (2002) modelling, especially in terms of the absence of economic theory to guide the specification and the complexity of the estimation strategy. Echoing Helberger and Knepel's (1988) critique, they also question the relevance of the causal and indicator variables that are employed. Giles and Tedds' specification and results are examined more closely by Breusch (2005a), who shows that the time path of their estimate for Canada has little to do with any underground activity, but mostly reflects price inflation and real growth in the observed economy. Moreover, the level of their estimate is a numerical accident with no connection to any evidence in the data.

My objective in the present paper is to look more broadly at MIMIC modelling as it is employed in this literature. A three-way distinction can be made between the method itself, the various ancillary treatments, such as data transformations and the post-model calibration that is called "benchmarking", and the modelling decisions

made when applying the method to a particular dataset. My starting point is to connect the method with the standard econometric models of linear regression and simultaneous equations. Much of the novelty in the MIMIC approach will be seen to reside in the labelling and interpretation of the calculations. The novel terminology and unfamiliar perspective are fostered by the adoption of specialist software packages such as LISREL® and Amos™. In most cases, exactly the same calculations can be described in terms that will be more familiar to the practising economist.

As examples of the method, I shall examine the three recent works mentioned above: Giles and Tedds (2002), Bajada and Schneider (2005) and Dell'Anno and Schneider (2003). In each case, I can replicate the MIMIC estimation results and the major inferences using both LISREL® and standard econometric software. There is considerable divergence in practice between the three applications, particularly in their interpretations of the latent variable and in their approaches to calibration and other adjustments. In every case, I discover undocumented transformations of the data, and speculate that the authors are unaware of making such transformations. As a result of these ancillary treatments, it is not always clear to the reader how, and by how much, the results of the MIMIC model are stretched and squeezed to fit some outside evidence.

I find instances where an inference about underground activity is sensitive to the units of measurement, so different substantive answers can be obtained simply by measuring the variables in different units. Sometimes this problem arises because of the form of calibration employed. In other cases, the dependence on units can be attributed to undocumented transformations of the data. Such sensitivity is an undesirable property in any measuring instrument, because the resulting measurement can be varied by changing a setting that is perceived to be irrelevant. The upshot is a method that lacks objectivity because it is open to manipulation and misrepresentation.

I examine critically the strategy of data differencing adopted in this literature to deal with unit roots and cointegration. The purpose of differencing is not always clear, but I show that the treatment is not an effective solution to any problem that matters, and may, in fact, cause serious problems. Independent of the issues of dynamic specification, the very idea of the underground economy as a latent variable is questionable. I provide evidence to show that the MIMIC model has precise statistical implications that are absent from this area of application.

In addition to the general principles examined in the main part of this paper, I have discovered many errors and anomalies while replicating the three studies. These additional findings are not essential to understanding the MIMIC method in general or its potential for estimating the underground economy, so they are gathered into an Appendix. However, this material does demonstrate some of the pitfalls that await users of the method, and contains important advice for readers who seek to interpret or employ the substantive results of the three studies.

Mimic and econometric models

The MIMIC model is described by Giles and Tedds (2002, Ch.6) as a relationship between a vector y (p x 1) of indicator variables and another vector x (q x 1) of causal variables. These are connected by an unobserved latent variable η (scalar) as follows:

$$y_t = \lambda \eta_t + \varepsilon_t \tag{1}$$

$$\eta_t = \gamma' x_t + \xi_t , \tag{2}$$

Where γ (q x 1) and λ (p x 1) are unknown parameter vectors. The error terms ε_t (p x 1) and ξ_t (scalar) are assumed to have zero means, variances $\Theta = diag(\theta_1,\ldots,\theta_p)$ and ψ, and to be uncorrelated with each other. A model consisting of (1) and (2) cannot determine the scale of all of the parameters, so a normalization condition is required. There are many possibilities, but Giles and Tedds adopt the convention of setting the first element of λ to be unity, as $\lambda_1 = 1$. The data are a time series of observations $t = 1,\ldots,N$. Estimation is typically by maximum likelihood, on the additional assumption that the error terms ε_t and ξ_t are jointly normally distributed and independent over time.

In the MIMIC model, x is weakly exogenous in the sense that all of these distributional statements are conditional on x. Thus, the model implies particular structures for the conditional mean and variance of the observed variables:

$$\mathrm{E}(y_t \,|\, x_t) = \mathrm{E}\left[\lambda(\gamma' x_t + \xi_t) + \varepsilon_t \,|\, x_t \right] = \lambda \gamma' x_t , \tag{3}$$

$$\mathrm{var}(y_t \,|\, x_t) = \mathrm{var}\left[\lambda(\gamma' x_t + \xi_t) + \varepsilon_t \,|\, x_t \right] = \mathrm{var}\left[\lambda \xi_t + \varepsilon_t \,|\, x_t \right] = \lambda \lambda' \psi + \Theta \tag{4}$$

These results can be written as a reduced form regression equation

$$y_t = \Pi x_t + v_t \tag{5}$$

where $\Pi = \lambda\gamma'$ and $v_t \sim (0,\Omega)$, and where $\Omega = \lambda\lambda'\psi + \Theta$. In general, the structure of the MIMIC model will imply restrictions on the reduced form parameters Π and Ω.

I wish to consider in more detail the case of two indicator variables, $p = 2$, since that is the nature of all three applications to be examined. In detail, then:

$$\lambda = \begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix} , \text{ so } \Pi = \lambda\gamma' = \begin{bmatrix} \gamma' \\ \lambda_2\gamma' \end{bmatrix} , \ \Theta = \begin{bmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{bmatrix} \tag{6}$$

It can be seen that the reduced form has $2q + 3$ parameters ($2q$ elements in Π and 3 more in Ω). However, the underlying model has $q + 4$ parameters (γ, λ_2, ψ, θ_1, θ_2). When $q>1$, as is typical, the reduced form will be restricted by the model. Writing out the restricted reduced form equations in full gives:

$$y_{1t} = \gamma' x_t + v_{1t} \tag{7}$$

$$y_{2t} = \lambda_2 \gamma' x_t + v_{2t} \tag{8}$$

where

$$\text{var}\begin{bmatrix} \\ v_{1t} \\ v_{2t} \end{bmatrix} = \begin{bmatrix} \psi + \theta_1 & \lambda_2 \psi \\ \lambda_2 \psi & \lambda_2^2 \psi + \theta_2 \end{bmatrix} \tag{9}$$

Here, the coefficient vector in the second equation (8) is in constant proportion to the coefficient vector in the first equation (7). There is no additional restriction on the variances in this case since, given λ_2, there are three distinct elements to the variance matrix and three parameters (ψ, θ_1, θ_2).[39]

It is useful to write out the structure in the standard econometric form of a simultaneous equation model for this leading case of $p = 2$ indicators. Multiply (7) by λ_2 and subtract the result from (8),

$$y_{2t} - \lambda_2 y_{1t} = v_{2t} - \lambda_2 v_{1t} = u_t \text{ (say)} \tag{10}$$

which gives the model as

$$y_{2t} = \lambda_2 y_{1t} + u_t \tag{11}$$

$$y_{1t} = \gamma' x_t + v_{1t} \tag{12}$$

where (12) is just a repeat of (7). This is formally identical to a two-equation linear simultaneous model, with two endogenous variables and q exogenous variables. In fact, maximum likelihood here defines the standard econometric procedure of limited information maximum likelihood (LIML) on equation (11), because the second equation is already in reduced form, and the covariance matrix between u_t and v_{1t} is unrestricted. The last point is seen here

$$\text{var}\begin{bmatrix} \\ u_t \\ v_{1t} \end{bmatrix} = \text{var}\begin{bmatrix} v_{2t} - \lambda_2 v_{1t} \\ v_{1t} \end{bmatrix} = \begin{bmatrix} \lambda_2^2 \theta_1 + \theta_2 & -\lambda_2 \theta_1 \\ -\lambda_2 \theta_1 & \psi + \theta_1 \end{bmatrix} \tag{13}$$

which has three free elements and determines three parameters, given that λ_2 is determined as a coefficient.[40] Thus, the indicator and causal variables of the MIMIC model match exactly the endogenous and exogenous variables of econometrics. The

[39] When there are more than two indicator variables in the model, so that $p > 2$, there are restrictions on the covariance matrix as well as among the coefficient vectors of the restricted reduced form.

[40] LIML recognises the structure of one equation of a system, while treating the other equations in their reduced form and ignoring any covariance matrix restrictions. It therefore uses the same information about the structure of the model as two-stage least squares, to which it is asymptotically equivalent. With more than two indicator variables in the MIMIC model, maximum likelihood estimation is not simply LIML, because the restrictions on the covariance matrix would be ignored in LIML.

measurement equations in the MIMIC model define the structural relationship by which the endogenous variables are jointly determined in the model.

While the MIMIC model can be interpreted and estimated as a standard econometric model of linear simultaneous equations, it retains its other interpretation in terms of the latent variable. The variance parameters in the MIMIC model can be solved from the econometric model as follows:

$$\theta_1 = -\text{cov}(u_t, v_{1t})/\lambda_2 \tag{14}$$

$$\theta_2 = \text{var}(u_t) - \lambda_2^2\theta_1 \tag{15}$$

$$\psi = \text{var}(v_{1t}) - \theta_1 \tag{16}$$

Since these are variances, all three must be positive for the estimate to be admissible as a MIMIC model, but LIML estimation will not impose non-negativity on the solution, and it is not difficult to construct examples in which any one of the implied variances is negative.[41] This is no different from the LISREL® and Amos™ packages, which by default do not restrict the variance estimates to being positive, although in some cases a warning message is issued when the estimated variance matrix is not positive definite.

One virtue of our interpretation of the MIMIC model as simultaneous equations is that it can be estimated without the specialist LISREL® or Amos™ software. Economists, who may be unfamiliar with such software and its conventions, can then see clearly what computations are being performed on their data. Often the simplest and most insightful way to apply LIML estimation is to recognise its equivalence to iterated generalized least squares (or Aitken) estimation for seemingly unrelated regression (see Pagan, 1979). This GLS procedure is available in many packages, such as the "sureg" command in Stata™. Of course, iterated GLS only yields directly the estimates of λ_2 and γ in (11) and (12), and perhaps the variance matrix in (13). Estimates of the other parameters in the MIMIC model can be recovered easily: the variances θ_1, θ_2 and ψ come from substituting the GLS variance and coefficient estimates into expressions (14)-(16).

The main use of the MIMIC model in this literature is to extract the latent variable η_t, which, in some sense, is interpreted as measuring the size of the underground economy. Since

$$E(\eta_t | x_t) = E(y_{1t} | x_t) = \gamma' x_t \tag{17}$$

the estimate of the latent variable is the predicted value of the first indicator variable (the one which is normalized to have unit coefficient in λ). Note that the prediction is made from the *restricted* reduced form, which will be estimated by LIML or GLS. The MIMIC model defines a proportionality relationship between the vectors of

[41] It can also be shown that at most one of the implied variances can be negative in this case of two indicator variables.

coefficients in the two reduced-form equations, so the prediction of the other indicator variable is just a rescaled version of the prediction of the indicator variable on which the normalization is made, where the factor of proportionality is the estimate of λ_2. By the invariance of maximum likelihood estimation, it makes no difference in principle which indicator variable is chosen for normalization, since the same estimates are defined, apart from the obvious change in scale. However, there are two important consequences of the normalization that should be considered: one is practical and the other may be important for interpreting the results.

In practice it is likely that one unrestricted reduced-form equation will fit the data much better than the other when estimated by OLS; in the language of instrumental variables, the exogenous variables may be much better instruments for one of the endogenous variables than the other. In that case, the restricted LIML estimates of the reduced-form coefficients will more closely resemble the unrestricted OLS estimates of the equation with the higher R-squared. Then the estimated latent variable will be similar to the unrestricted OLS prediction from the better-fitting reduced-form equation, perhaps scaled by λ_2 if it is necessary to normalize on the other indicator variable. As a practical matter in estimation, if the reduced-form equations have very different fits by OLS, the iterations will be found to converge faster and more reliably if the model is normalized on the indicator (endogenous) variable with the higher R-squared.[42]

Such practical considerations aside, the question of how to normalize the model is usually seen as a matter of convention and convenience, but it may affect interpretation of the results. In the standard assumption of $\lambda_1 = 1$ for the model of equations (1) and (2), the latent variable is linked to the first-listed indicator variable by the normalization. Reordering the variables will switch another variable to become the normalizing indicator, and hence it will rescale the latent variable. Thus, there is a degree of indeterminacy in scale, which needs to be resolved if the latent variable is to be interpreted as an estimate of the underground economy.

In recognition of this ambiguity, the latent variable is sometimes called an "index". The approach in the literature is to set the absolute level of the estimate by requiring the index to pass through a particular value at a particular time, in a step that is often called "benchmarking" but is more accurately described as calibration. This is analogous to the familiar treatment of an index of prices, in which the series is set in a base period to an arbitrary value of one or 100, and the rest of the series is scaled accordingly. In the present case, the benchmark is not some arbitrary number, but rather is found from other modelling carried out independently of the MIMIC model. If the calibration is multiplicative, it will preserve the proportional relationships in the series (as with a price index). In such a rescaling operation, it will make no difference to the final inference which of the indicator variables is

[42] This is similar to recent findings in the "weak instruments" literature; for example Hamilton, Zha, and Waggoner (2007).

used for normalization. However, as we shall see, the calibration is not always done in this way and, as a result, the inference is not always invariant to the normalization.

Three applications to the underground economy

I shall present three applications in which MIMIC modelling is used to estimate the underground economy: Giles and Tedds (2002), Bajada and Schneider (2005) and Dell'Anno and Schneider (2003). There is much that is common to these studies, and the latter two papers cite the earlier as a forerunner, but I also find considerable variety among the applications in their approach and interpretation. Unfortunately, the reader is not always informed of these differences by the documentation provided. There are instances in all three works where the description of a procedure, or the context of references to other literature, suggests one approach when, in fact, a different calculation is needed to obtain the stated results. So, while the explanations of *why* something is done are drawn from the papers themselves, I rely on my own careful replications of the calculations to determine *what* is actually done to the data.[43] These replications employ the original data or a close facsimile of them.[44]

This section considers only those aspects of modelling and reporting that are essential to understanding the various ways in which MIMIC modelling is used. I shall focus on the issues of specifying and estimating the model, calibrating the index, and interpreting the resulting time series. Other errors and anomalies uncovered in the process of replicating the three studies are described in the Appendix. This additional information will be useful for readers who seek to understand the substantive results in the three studies.

To simplify the discussion, I shall define a standard notation for the common variables. In Table 1, variables with names of one and two characters appear in at least two of the studies, or are components of constructed variables, while those variables with longer names are used uniquely. I shall use the abbreviated, symbolic names, even when the original study might use a longer description, for example using $\Delta 1n(YD/(P \times N))$ rather than "the proportional growth rate of real, per capita, disposable income". Some fine distinctions may be obscured by this practice (such as the units of measurement or the base year for a price index), but such subtleties can be recovered when they are needed.

[43] Replication is valuable as a springboard to new inquiry from existing published research, and is an efficient method of purging incorrect results from the body of accumulated knowledge. See McCullough, McGeary, and Harrison (2006) for an evaluation of replication in applied economics and analysis of the data archives of the *Journal of Money, Credit and Banking*. Anderson, Greene, McCullough, and Vinod (2008) conduct a similar investigation at the Federal Reserve Bank of St Louis.

[44] I thank Lindsay Tedds for supplying the Canadian data used in Giles and Tedds (2002), and Christopher Bajada for the Australian data from Bajada and Schneider (2005). The Italian data described in Appendix 1 of Dell'Anno and Schneider (2003) are taken from *OECD Economic Outlook* and the Bank of Italy's online database.

Table 1. Definitions of variables

Y	....	nominal observed GDP
C	....	currency held by public
YD	....	nominal disposable income (= $Y - TH - TB + W$)
TH, TB, TI		taxes collected from households and business, and indirect taxes
W	....	welfare state benefits and transfers
P	....	price level
L	....	labour force
M	....	unemployment rate
N	....	national population
S	....	number of self-employed persons
U	....	nominal underground income
MULT	....	number of male holders of multiple jobs
SELF	....	nominal incomes of self-employed persons
ERTE	....	nominal $Can/$US exchange rate

All three applications employ two indicator variables (the vector of *y*s) and a short list of causal variables (the vector of *x*s). The indicator and causal variables for each study are listed in Table 2, along with other summary information that will be discussed in detail under the individual studies. The pair of indicators in each case consists of observed GDP in some measure (real, or real per capita, in a logarithmic transformation) and currency holdings by the public in some similar measure. The causal variables are more varied, but typically include a range of tax rates and some measures of real disposable income per capita, the level of employment or unemployment, the extent of self-employment, and welfare state transfers or total government spending. In all three cases there is some sequential differencing of the variables before the model is fitted, as a treatment for unit roots and cointegration, although there are some differences in the criteria used to make decisions about the differencing. There is also divergence among the applications in the extent to which they standardize the means and standard deviations of the variables before estimation. Further differences will be observed between the three studies in their interpretations of the latent variable and, in particular, in the various ways they calibrate the index after estimation.

Study 1: Giles and Tedds (2002)

Before the MIMIC model is estimated, the variables in this study are differenced to the extent that secures their stationarity, according to the results of individual unit root tests. *C* and *SELF* are differenced twice, and most of the other variables are differenced once, while *YD*/(*P* x *L*) is not differenced at all. The differenced variables are then all transformed into deviations from means and scaled to have unit standard deviation. These last two data operations are not mentioned at all in the published documentation, which is surprising because both are unusual in econometrics. Perhaps these transformations have been made unintentionally, most

likely by accidentally invoking an option in the estimation software.[45] Later sections will explore the consequences for inferences about the underground economy of the (documented) differencing operations and the (undocumented) transformations of location and scale in the variables.

Table 2. Summaries of the three studies

Study 1: Giles & Tedds (2002), Model 6
Indicators: 1n(Y/P)†, C **Causes:** *MULT*, *SELF*, *YD/*(*P x L*), *ERTE*, *TB/Y*, *TI/Y*, *M* **Data:** Canada, annual 1976-1995 **Specification:** in levels **Differencing:** levels or first differences or second differences **Undocumented:** deviations-from-means and unit standard deviation (standardized) **Index:** 100U/Y in percentages **Calibration:** multiplicatively, to a level of 9.45 per cent in 1986 **Base for levels:** set by the calibration benchmark.
Study 2: Bajada & Schneider (2005)
Indicators: 1n($Y/$(*P x N*)), 1n($C/$(*P x N*))† **Causes:** 1n(*YD/*(*P x N*)), 1n(*TH/Y*), 1n(*TB/Y*), 1n(*TI/Y*), 1n(*W/YD*) **Data:** Australia, quarterly 1966q2 to 2003q3, deseasonalized **Specification:** in differences **Differencing:** first differences **Undocumented:** deviations-from-means **Index:** 100Δ1n(U/Y), integrated and transformed to 100U/Y in percentages **Calibration:** additively, to a growth rate of 0.0021 per cent in 1980q2 **Base for levels:** approximately 13.5 per cent in 1968q2?
Study 3: Dell'Anno & Schneider (2003), Model 3-1-2b
Indicators: 1n(Y/P)†, 1n(C) **Causes:** ($TH + TB + TI$)/Y, G/Y, S/L **Data:** Italy, semi-annual, 1960s1 to 2000s2 **Specification:** in differences **Differencing:** first differences, causes and income indicator multiplied by 100 **Undocumented:** deviations-from-means **Index:** $\Delta U/P$, integrated to U/P in units of 10 billion euros **Calibration:** none **Base for levels:** 19.7 per cent in 1978s2.

$\dagger$ = Normalization on this variable

We can interpret Giles and Tedds as specifying the model in the original levels variables, although they estimate the model after variously differencing the variables. This interpretation follows from the way they form the latent variable or index and how they subsequently calibrate the index to become their estimate of the underground economy. In this study, the vector y_t contains the two indicator variables 1n(Y/P) and C, and the vector x_t contains the seven causal variables, *MULT*, *SELF*, etc. We can write the indicators in the estimation model as the vector $\tilde{y}_t$, which contains Δ1n(Y/P) and $\Delta^2 C$, after these variables have been transformed to deviation-from-means and scaled to have unit standard deviation. Similarly, we can represent the causes in the estimation model as the vector $\tilde{x}_t$, which contains

[45] Tedds and Giles (2005) deny that the variables used in Giles and Tedds (2002) are standardized. However, the estimation results can be replicated if, and only if, the variables are transformed in this way.

$\Delta MULT$, $\Delta^2 SELF$, etc., after they have been transformed by location and scale in the same way. Thus, the model is specified just as it is written in equations (1) and (2) with the variables y_t and x_t, but the maximum likelihood estimator is applied after these variables are replaced by $\tilde{y}_t$ and $\tilde{x}_t$. The index in Giles and Tedds, however, is not calculated as $\hat{\eta}_t = \hat{\gamma}' \tilde{x}_t$, which is the direct estimate of the latent variable from the estimation model, but rather as $\hat{\eta}_t = \hat{\gamma}' x_t$, which applies the estimated coefficients to the original, untransformed, causal variables. It is this latter form of index that is scaled in Giles and Tedds' calibration operation, on the grounds that the scale of the index is indeterminate in MIMIC modelling. Clearly, then, they interpret the MIMIC *model* on the original data, even though the *estimates* are derived by fitting the model to transformed data.

Calibration or "benchmarking" in Giles and Tedds is done from a separate currency demand model fitted to data similar to the MIMIC model. From this auxiliary model, an estimate of the underground economy at 9.45 per cent of official GDP is derived for 1986. The index from the MIMIC model is then set to this benchmark, and the rest of the estimated series is found proportionally:

$$ug_t = 9.45 \times \hat{\eta}_t / \hat{\eta}_{1986} \quad \text{for } t = 1976, ..., 1995 \tag{18}$$

While this formula is not stated explicitly in Giles and Tedds (2002), it is described in words in Giles (1999) and its use by Giles and Tedds is confirmed by replication of their results. It is simply a scaling operation, so it preserves the proportional relationships between the measurements in different years:

$$ug_t / ug_s = \hat{\eta}_t / \hat{\eta}_s \quad \text{for all } t \text{ and } s \tag{19}$$

Thus, the calibrated series will be the same, whichever of the indicator variables is used for normalization, because the arbitrary choice of scale that is imposed by the normalization is removed in the calibration operation.

The scaled series ug_t is interpreted by Giles and Tedds as estimating the underground economy income in Canada as a percentage of observed GDP, that is $100U/Y$. Their resulting estimate is a 20-year time series that grows from a low of 3.46 per cent of GDP in 1976 to a high of 15.64 per cent in 1995, passing through the benchmark value of 9.45 in 1986.[46]

Because of the multiplicative scaling in (18), the overall *level* of this estimate of the underground economy is derived from the benchmark value, which comes from the separate currency demand model. On the other hand, the *time path* of the estimate is due entirely to the MIMIC model. The series is 4.5 times higher at the end of the 20 years than at the beginning, which is equivalent to a compound rate of increase of 7.8 per cent per year. This phenomenal growth rate is more remarkable for being relative to observed GDP, which in real terms grows by 64 per cent in the same period. Thus, according to this estimate, the level of underground income in

[46] The results are shown in Giles and Tedds (2002, Table 7.1 and Figure 7.2).

Canada, in real dollars using the implicit GDP price deflator, increases more than seven times in 20 years. At the same time, the observed economy much less than doubles in size. This astounding growth is the main inference from the MIMIC model.[47]

Study 2: Bajada & Schneider (2005)

Although this study refers to Giles and Tedds as a progenitor, the approach here is very different. Each variable is differenced only once, so the indicators and causes in the estimation model are all quarterly proportional growth rates of the underlying economic variables. This uniform single differencing may be a matter of luck, because the only discussion of the strategy is that "the data used in the MIMIC estimation were differenced after testing for the presence of a unit root" (p.394). However, there is also consistency in the way the variables are uniformly in logarithmic form and are either major economic aggregates measured in real terms per capita, or tax and welfare payments in proportion to an aggregate of income. The variables are all calculated as deviations from means in the estimation model (although that transformation is not documented), but there is no scaling of the variables to have unit standard deviation, as there is in Giles and Tedds.

In this case, it is appropriate to think of the model as being specified *and* estimated in quarterly growth rates. The authors interpret the latent variable in the estimation model as the (percentage) growth rate of the ratio of underground income to observed GDP, $100\Delta 1n(U/Y)$. This quantity is first calibrated, and then integrated up from the growth rates to form an index of the level of $100U/Y$. A second round of adjustment is employed later to allow the level of the underground economy to be inferred from an estimate of its growth rate.

Again, calibration is done from a currency model estimated from similar data to the MIMIC model.[48] As Bajada and Schneider say, "A quarterly growth rate was chosen from the results of the currency-demand model as a benchmark to produce a growth rate of the underground economy implied by the MIMIC index" (p.394). Although the authors do not specify how this operation is done, from replication of their results it is apparent that the calibration is not the *multiplicative* adjustment of Giles and Tedds, but instead a novel form of *additive* adjustment. Suppose we write the latent variable derived by the prediction formula (17) with estimated coefficients as $\hat{\eta}_t = \hat{\gamma}'\tilde{x}_t$. Here, $\tilde{x}_t$ contains the causal variables of the estimation model, in this case $\Delta 1n(YD/(P \times N))$, $\Delta 1n(TH/Y)$, etc., each adjusted to deviations from means. Then, the operation used for calibration by Bajada and Schneider can be written as

[47] See Appendix for a further discussion of the modelling and these results.

[48] Calibration in Bajada and Schneider is done from a slightly modified form of the currency demand model of Bajada (1999). See Appendix for a discussion of the currency model and the calibration results.

$$ugd_t = ugd_0 + \hat{\eta}_t - \hat{\eta}_0 \text{ for } t = 1,...,N \quad (20)$$

Where ugd_0 is the benchmark value of the series of differences, taken from the currency model in the benchmark period $t = 0$, and $\hat{\eta}_0$ is the value of the latent variable from the MIMIC model in the same period. The magnitude of ugd_0 and the timing of the benchmark period are unstated by the authors, but appear to be set at 0.0021 in 1980q2.[49] This procedure simply matches the growth rate from the MIMIC model to that of the currency model in the benchmark period, by adding a constant to the growth rate each period. In contrast to Giles and Tedds, who *scale* the predicted latent variable in the levels model to meet the benchmark, the procedure adopted here is to *slide* the latent variable in the differences model into place against the benchmark.

Bajada and Schneider do not offer any rationale for this additive form of post-estimation adjustment. It certainly does not satisfy the principle emphasised by Giles and Tedds that the scale of the latent variable from a MIMIC model is arbitrary and must be fixed on other information. Here, it is the level of the latent variable that is being adjusted, not its scale. Nor is this form of modelling invariant to the choice of normalizing indicator variable. Normalizing on the income variable rather than currency will change the scale of the coefficients, and hence that of the index. The new scale does not cancel in this procedure, as it does in Giles and Tedds' multiplicative calibration; so in this case, with additive calibration, the choice of indicator for normalization is substantive and not just a mathematical or computational convenience. It might therefore be supposed that Bajada and Schneider attach some meaning to the normalization they adopt. However, all we are told is that "the coefficient on currency holdings is constrained to +1.00 in order to identify the system and make the parameter estimates more easily comparable with one another" (p.393).

It is tempting to suppose that a change of scale in the original variables is equivalent to an additive shift in the logarithms of those variables, but here a constant is added to the growth rates, which becomes an additive linear time trend in the levels of the logarithms, and hence a multiplicative exponential trend in the underlying economic variables when the logarithm transformation is reversed. There is no dimension is which this procedure is an adjustment to fix an unidentified scale. Curiously, however, there is apparently one small virtue of this form of calibration. It turns out not to matter whether the undocumented deviations-from-means transformation of the estimation variables is included or ignored when the latent variable is calculated by the formula $\hat{\eta}_t = \hat{\gamma}' \tilde{x}_t$. The difference between the two approaches will be an additive constant, which will then cancel when the index is adjusted additively to its benchmark.

[49] This specification of the benchmark is deduced from inspection of Bajada and Schneider (2005, Figure 4).

Bajada and Schneider interpret the calibrated series called ugd_t in (20) as the percentage growth rate of the underground economy relative to official GDP. The growth rate is then integrated to get the levels: "Using the currency demand approach to benchmark the starting values of the shadow economy, the MIMIC index was used to generate the level path (as a percentage of GDP) for the shadow economy" (p.395). Unfortunately, no details of this second round of adjustment are provided, and I have been unable to reconstruct precisely the method that was used. In any case, it is misleading to call this second stage calibration, much less benchmarking, because it does not adjust the measuring device against external data; it simply fixes a base point that converts a series of growth rates into a series of levels. Perhaps *anchoring* is a better term.

Taking the results of the second adjustment operation at face value, the level of the underground economy is shown in Table 3 of Bajada and Schneider to hover close to 13.9 per cent of recorded GDP for the period 1993-2003.[50] The annual figure over this decade never moves more than 0.3 of a percentage point from its average. This is a remarkably flat time series by any comparison, both with the estimates for Australia by the method of currency demand modelling in Bajada (1999) and with results reported for other countries. However, since outside information is used to fix both the growth rate of the index (by calibration) and its level (by anchoring), there is not much in this result that can be attributed to the MIMIC model.[51]

Study 3: Dell'Anno & Schneider (2003)

Dell'Anno and Schneider also cite Giles and Tedds as a forerunner, but they employ a methodology that differs in certain crucial ways from both that and the study by Bajada and Schneider. The variables are uniformly first differenced for estimation, apparently without prior testing but rather on the principle that, "In order to eliminate the non-stationarity of the variables, the [causes] are taken as first differences, [while the indicators] are converted in the first differences of logarithm" (p.102, fn.13).[52] Both indicator variables, when differenced, are interpreted as semi-annual growth rates, of real income and nominal currency holdings, respectively. In another parallel to Bajada and Schneider, the causes are taxes or government expenditures in proportion to GDP, or labour force categories in proportion to the total. The variables are all transformed to deviations from means for estimation (again undocumented), although there is no scaling to unit standard deviation.

[50] The final estimates are taken "at face value" because I cannot replicate them. The units of measurement are muddled and there are obvious contradictions between the growth rates in Figure 4 of Bajada and Schneider and the levels in their Table 3. See Appendix for details.

[51] Further implications of fixing both the growth rate and the level of the index are pursued in the Appendix.

[52] There are unit root tests in a sole-authored discussion paper by Dell'Anno (2003), which appears to be an earlier version of the Dell'Anno and Schneider paper.

Also in common with Bajada and Schneider, the model is specified *and* estimated in first differences. However, in this case the authors interpret the latent variable in the estimation model quite differently – as the change in real underground income, $\Delta(U/P)$. This differs in dimension from both indicator variables, which in their differenced form are growth rates of the underlying economic variables. In further contrast to the other studies, the latent variable from the estimation model is not calibrated to an outside estimate, but instead is assumed (implicitly) to be measured in units of 10 billion euros. This quantity is then integrated up from the changes to form an index in the levels of U/P. An external value from other studies is used to fix the overall level of the series to a value of 19.7 per cent in 1978s2. As noted earlier, this is not calibrating the MIMIC index in the manner of Giles and Tedds, but rather anchoring the time path of the levels, which is required because the levels are being constructed from estimates of the changes.

The index is not obtained by Dell'Anno and Schneider directly from the estimation model, but rather is constructed as a separate predictor $\hat{\eta}_t = \hat{\gamma}'\check{x}_t$, where $\check{x}_t$ contains the differenced causal variables. The distinction here is that the deviations-from-means transformation applied to the data to obtain the parameter estimates in the MIMIC model is ignored in forming $\check{x}_t$. There is no additive calibration adjustment here, as there is in Bajada and Schneider, so the two ways of forming the index will differ by a constant. Since this index is being interpreted as the change in real underground income, the constant difference will affect every point in the final series (except the one point where it is anchored on external information).

With no calibration of the latent variable obtained from the MIMIC model, this application does not conform to the principle that the scale of the index is arbitrary and must be fixed on other information. The inferences in this case will depend materially on the choice of indicator variable used for normalization. There are suggestions that the authors are troubled by the contradictions that arise. On the one hand, they recognize that their choice of normalizing indicator $\ln(Y/P)$ is material: "this variable ... is chosen as *variable of scale* (or *reference variable*)" (p.105, emphasis in original), and "The choice of the 'sign' of the coefficient of scale (λ_{11}) is based on theoretical and empirical arguments" (p.106).

However, elsewhere they accept that normalization should be a matter of convention and convenience:

> ...in order to estimate not only the relative size of the parameters but their levels, is necessary to fix a scale for the unobserved variable. A natural normalization would be to assign a unit variance to the latent variable but a more convenient alternative is fix a non-zero coefficient to reduced form (p.105, fn.19).
>
> The value of the fix parameter is arbitrary, but using a positive (or negative) unit value is easier to find out the relative magnitude of the other indicator variables (p.106).

To further confuse the issue of normalization, the authors do not simply choose which of the indicators 1n(Y/P) and 1n(C) is given a unit coefficient; they specify that 1n(Y/P) should have a coefficient of *negative* one. The "theoretical and empirical arguments" for this decision are not made explicit, but it seems that the objective is to ensure that key coefficients in the structural equation for the latent variable have the desired sign. If the normalizing were done in the usual way, the inference would be the unfortunate one that higher growth in the tax burden, or in the size of government, or in the extent of self-employment, all lead to *reductions* in the size of the underground economy.

Given that normalization is arbitrary, in both magnitude and in sign, it is impossible to infer any relationship between the latent variable that represents the underground economy and the endogenous variable on which the normalization is made. However, the authors feel no such inhibition in claiming: "In our analysis, we find evidences to support the hypothesis of negative relation between Italian shadow economy and official growth rate of GDP" (p.106) and "The relationship between underground economy and growth rate of GDP (Y1) is negative" (p.112).

The final output of this study is a time path of underground income in proportion to official GDP ranging from over 40 per cent in the early 1960s, down to 15 per cent in 1975-77, and then back to around 25 per cent in 2000. Along the way, it passes through the anchored value of 19.7 per cent in 1978 (Dell'Anno & Schneider, 2003, pp.110-111, Figures 2 and 3). The shape of the path depends on the twin assumptions that the index from the MIMIC model is measuring changes in real underground income and the measurement is in units of 10 billion euros. Any other interpretation will give a materially different time path, although both parts of the assumption are quite arbitrary (and unstated in the paper). The choice of the income variable for normalization and the transformation of the variables in the estimation model to deviations from means both influence the result – although the former is arbitrary and the latter undocumented. Furthermore, the assignment of a negative coefficient to the normalizing indicator variable will reverse the sign of the latent variable. Since the latent variable is interpreted as a series of changes, that decision will *invert* the time path of the final result.

Deviations from mean and unit standard deviation

I have been able to replicate the estimation results of these three studies without using the specialist LISREL® or Amos™ software. This independent reconstruction of the estimates reveals that the variables have been transformed to deviations from means, and in one case also scaled to have unit standard deviations, although these transformations are undocumented. The finding that all three studies make at least one of these transformations, apparently without the authors being aware of doing so, is at once puzzling and alarming. I shall examine the nature of these transformations and explore their effects on the inference that is made of an underground economy. In every case, the transformation applied in estimation is

ignored when the prediction is formed, with the result that the inference acquires some very undesirable properties. I also speculate on how such undocumented transformations might have occurred.

Consider a simple linear regression model between a scalar y and a vector x (q x 1)

$$y_t = \gamma_0 + \gamma_1' x_t + \varepsilon_t \tag{22}$$

where the intercept scalar γ_0 and slopes vector γ_1 (q x 1) are unknown parameters. The error term ε_t is assumed to have zero mean and constant variance, and to be serially uncorrelated for observations $t = 1,\ldots,N$. In a well-known set of results, the least squares estimates are:

$$g_1 = \left[\sum x_t^* (x_t^*)'\right]^{-1} \sum x_t^* y_t^* \text{ and } g_0 = \bar{y} - g_1' \bar{x} \tag{23}$$

Here, the variables x_t^* and y_t^* are transformations of the original variables into deviations from their sample means:

$$x_t^* = x_t - \bar{x} \text{ where } \bar{x} = N^{-1}\sum x_t \text{, and similarly for } y_t^* \tag{24}$$

The original model can be written as

$$y_t^* = \gamma_1' x_t^* + \varepsilon_t^* \tag{25}$$

which has transformed variables but no intercept. The first equation in (23) indicates that least squares on (25) gives the same slope estimate as the original model (22). The second equation in (23) shows how to extract the implied intercept.

While estimation is the same in both transformed and untransformed variables (provided an intercept is fitted in the latter case), more care is needed when making predictions. For one thing, the models have different dependent variables, so the targets of prediction are different. Using the standard form of the predictor in both cases, $g'_1 x^*_t$ predicts y^*_t in the transformed model, while $g_0 + g'_1 x_t$ predicts y_t in the original model, in both cases giving an unbiased prediction. However, when a model is fitted to variables that have been transformed to deviations from means, but that transformation is ignored when the predictions are formed, the result will be a hybrid predictor of the form $g'_1 x_t$. This makes a biased prediction of *both* y^*_t and y_t. What is more, the bias depends on the intercept in the model, so if any variable in the equation is in logarithmic form, the intercept will change with the units of measurement of that variable, making the whole procedure sensitive to the change in units. This is a clear deficiency in what seems to be common practice in forming the latent variable after MIMIC estimation.[53]

[53] The correct predictor in the transformed model, $g'_1 x^*_t$, has the property that it is zero on average. If this predictor is interpreted as a series of changes or growth rates and integrated to form an index for the levels, the resulting index has the property that its net change over the estimation period is zero. This will imply that the estimated underground economy is the same size at both ends of the

Another transformation that is sometimes considered in linear regression is to write the model with each variable standardized by subtracting its mean and dividing by its standard deviation

$$y_t^{**} = \beta_1' x_t^{**} + \varepsilon_t^* \tag{26}$$

where the standardized variables

$$y_t^{**} = (y_t - \bar{y})/s_y \text{ where } s_y^2 = N^{-1}\sum(y_t - \bar{y})^2 \text{, and similarly for } x_t^{**}$$

have mean of zero and standard deviation of one. The estimates of coefficients in (26) by least squares are called various names; for example, they are "normalized beta coefficients" in Stata™ and "standardized beta coefficients" or just "betas" in SPSS. The connection with the usual estimates is:

$$b_j = (s_x/s_y)\, g_j \text{ for } j = 1,...,q \tag{27}$$

Standardized betas are occasionally used to make statements about the relative importance of the independent variables in a multiple regression model. They are invariant to the units in which the variables are measured so, if there is a change of units that rescales one or more of the variables, the standardized betas are unaffected. Again, care is required in making predictions from the transformed model that the transformed predictor variables are used and that the object of prediction is the transformed dependent variable. Otherwise, as we shall see below, the prediction is not only biased, but is also sensitive to the scale of the units in which the variables are measured.

It is natural in the approach taken by the software packages LISREL® and Amos™ to think of the data being first transformed to deviations from means, and sometimes also transformed to unit standard deviation. The statistical orientation of the user community tends towards multivariate analysis and the use of factor structures to represent patterns of covariance and correlation. The language and assumptions of the software reflect that orientation. Hence, the structural model of Section 2 above might be described as a problem of summarizing the covariances of the data contained in the extended vector $z = (y',x')'$, using a conditional mean with the structure $E(y|x) = \lambda\gamma' x$, a conditional variance $\text{var}(y|x) = \lambda\lambda'\psi + \Theta$, where Θ is diagonal, and without restricting the covariance matrix of x. Given the focus on modelling covariance in this approach, it is often assumed that the means of the variables have already been removed. Hence, the default setting in the software is to subtract the means from each of the variables before fitting the model, thus transforming it in this way. For example, unless there is an "MA" instruction on the "DA" line in the input file, LISREL® will automatically transform the data to deviations from means. This should not surprise the economist: subtracting the

period. None of the three applications actually does this: they either benchmark the differences before integrating or they use a different (and incorrect) prediction formula.

means is equivalent to fitting an intercept in a linear regression, which is also the default in most econometric software.

The prior transformation of all variables to have unit standard deviations is also quite natural in this setting. It corresponds with a focus in the analysis on modelling the correlations of the data rather than the covariances. If the model being fitted is one like simple factor analysis that can be described entirely as restrictions on the correlation structure of the data, then it may be convenient to transform in this way. Indeed, LISREL® has options to input the data in the form of a correlation matrix if that is convenient to the researcher. When the data are input as variables, not correlations, there are options that include transforming the variables in the estimation model to have unit standard deviation. Again, in LISREL®, the "SC" option on the "OU" line will give a fully standardized solution. There are equivalent options in other software: for example, in Stata™ the option "beta" on the "regress" command will output the standardized regression coefficients.

As we have noted, a faulty predictor of the latent variable will be employed when the researcher is unaware that the model is estimated on transformed variables. In Giles and Tedds, the estimation variables are fully standardized (transformed to deviations from means and adjusted to unit standard deviation), so the coefficient estimates are invariant to any changes in the units of measurement of the variables. For example, the variable *SELF* is measured in their data file in units of thousands of dollars a year. If all the values of the variable were divided by a thousand or a million, so the new units of measurement become millions of dollars or billions of dollars a year, exactly the same coefficients would be obtained in the MIMIC model because of the standardizing transformation. However, the predictor of the latent variable is formed by applying these standardized coefficients to the original variables. This hybrid form of predictor is not only biased, but is also sensitive to the units in which the variables are measured.

In the case of ordinary regression coefficients, any rescaling of a variable is compensated by an inverse scaling of its coefficient, so the product of the two remains invariant when a predictor is formed by linear combination; but when standardized coefficients are applied to non-standardized variables, no such compensation will occur. The coefficient remains constant as the variable is rescaled, so the product of the two elements changes with the scale of the variable. With more than one causal variable in the model, this will not be simply a scaling of the predictor (which might be removed subsequently by multiplicative calibration), but a more complicated set of changes to the relative weights of the variables in the linear combination. Thus, the final inference will be altered materially by the choice of units.

There are further problems with the hybrid predictor as used by Giles and Tedds. When standardized coefficients are applied to variables that are measured on vastly different scales, one or two of the variables are likely to dominate in the linear combination that forms the predictor. In the Giles and Tedds case, it turns out that

just one causal variable dominates the latent variable, and hence contributes almost all of the movement over time in the index of the underground economy. That variable is *SELF*, the nominal incomes of self-employed persons, measured in thousands of dollars a year. None of the more plausible variables in their model, such as the various tax rates, has any effect on their estimate.[54]

Bajada and Schneider employ the deviations-from-means but not the unit standard deviation transformation. In principle, the hybrid prediction strategy of applying the coefficients from the transformed model to the original variables will yield a biased predictor in this case. Also, the hybrid predictor will be sensitive to the units of measurement of any of the variables, which are all in logarithmic form. Happily, as we saw earlier, the additive form of calibration they use in forming the index will compensate for the form of the predictor. There remains the issue of a model that is incorrectly described, because the transformation is not reported, nor is the implicit intercept noted. We also observe that the unusual form of calibration in this application imposes an arbitrary solution to the identification problem in the MIMIC model. The results of this study would be substantively different if another, equally arbitrary, normalization of the latent variable were adopted.

In Dell'Anno and Schneider, the data are similarly transformed to deviations from means but not to unit standard deviations. The same criticism applies in this case of a model that is inadequately described, having either undocumented data transformations or a missing intercept parameter. As in the other applications, prediction of the latent variable is biased, because the means of the variables are removed for estimation but included when forming the predictor. Also, the construction of the index is sensitive to the units of measurement of the indicator variables, which both appear in logarithmic form. However, all of these are minor quibbles in the face of the larger problem noted earlier – that the units of the resulting index in this study are simply invented!

Differencing and co-integration

The aggregate time-series data used in all of these studies typically contain trends that may be attributed to unit roots. The reaction in all cases is similar:

> Before one can use the data ... appropriately to estimate models of the form given by [equations (1) and (2) above], one must check for the presence of unit roots. ... [W]e differenced the various data series appropriately to make them stationary. We then used them in this "filtered" form to estimate the MIMIC models... Usually, rather than proceeding directly to modelling after the unit root tests, one would also consider the possibility of cointegration. Unfortunately, there is

[54] See Appendix for more detail and references.

> no established literature to serve as a guide to this procedure in the context of MIMIC models (Giles & Tedds, 2002, p.128).

Dell'Anno and Schneider quote the final two sentences of the above passage and add, "... in some cases, to eliminate the non-stationarity in the time series, the variables are transformed (first differences and growth rates)" (p.107, fn.22). Bajada and Schneider are less informative about their motives and criteria, and simply say "... the data used in the MIMIC estimation were differenced after testing for the presence of a unit root" (p.394).

It is not entirely clear why unit roots are considered to be a problem in this setting. Somewhat earlier in their book than the passage quoted above, Giles and Tedds suggest one issue:

> Essentially, the point is that before one estimates a MIMIC model one must establish the properties of the data; otherwise, the result may be estimates that have undesirable statistical properties and hence measures of the latent variable that are meaningless (Giles & Tedds, 2002, p.104).

A different motive is indicated when these authors later seek to clarify their method:

> It is generally accepted that when modeling with time-series data, these data must first be tested for the presence of unit roots; if these are detected (and in the absence of cointegration), they are rendered stationary in order to avoid the consequences of estimating spurious regressions. That is, the model's coefficients are obtained using the stationary series, but the model's predicted values are calculated using the original data (Tedds & Giles, 2005, p.395).

Thus, two distinct dangers are identified: a meaningless latent variable because the coefficient estimates on which it is formed have undesirable statistical properties, and the risk of estimating spurious relationships.[55]

As I shall show, the act of differencing the variables before fitting the MIMIC model cannot solve the first of these supposed problems, while the second is simply irrelevant to the task at hand. Either the model is a relationship in the levels, in which case differencing is mildly or seriously damaging, or it is a relationship only in the differences, in which case there is no justification for forming an index in the levels. I shall consider both of these possibilities in turn.

On the first hypothesis, consider a model in the original levels of the variables. If the model consisting of equations (1) and (2), together with the assumptions on the variances and covariances of the errors, is a correct description of the process generating the data, there is no reason for concern about unit roots and cointegration. In that case, if the variables have unit roots they must be cointegrated,

[55] The term spurious *regression* seems curious here, since the MIMIC model supposedly represents a set of structural relationships, not simply statistical regression.

with two distinct cointegrating vectors since (7) and (8) describe linear combinations of the variables that are stationary (in fact, the linear combinations are white noise). There is a particular relationship between the two cointegrating vectors in this case, which follows from the structure of the MIMIC model. Of course, the conventional asymptotic distribution theory may not apply to the coefficient estimates, because the exogenous (causal) variables *x* will not have moments that converge in the way that is usually assumed in applications of maximum likelihood to independent data; but the coefficient estimates will be consistent, so the predictor will be cointegrated with each of the endogenous (indicator) variables. As in the standard theory, the latent variable is the fully efficient predictor of the normalizing indicator variable.

Estimating the model after differencing the variables either throws away information relative to fitting the model in the levels of the variables, or it imposes incorrect assumptions on the model. At best, the strategy leads to an efficiency loss, although there may be more serious consequences. On the one hand, provided the coefficients are consistent estimates (they may not be so, see below), the index formed from these estimates and the variables in levels will be cointegrated with the indicator variables. The asymptotic theory indicates that the estimates from the model in levels will be "super consistent" in the sense that they converge to the true parameter values at a much faster rate than the conventional root-*N* consistency. Thus, the variances of the coefficients in the two approaches may be of different orders of magnitude even in moderately-sized samples. So, while the cost in this case is only inefficiency in the coefficient estimates that arises from needless differencing, such losses may indeed be large. On the other hand, if the coefficients estimated after differencing are not consistent, the latent variable will not be cointegrated with the indicator variables. In that case, the outcome will be a predictor that has no long-run relationship with the endogenous variables it is supposed to predict, which is not a satisfactory solution to the problem of unit roots.

Differencing will return consistent estimates when the model satisfies all the assumptions in the levels variables, provided the *same* degree of differencing is applied throughout. To see this in a single-equation example, suppose the model is

$$y_t = \gamma_1 x_{1t} + \gamma_2 x_{2t} + \varepsilon_t \tag{28}$$

where ε_t is white noise and uncorrelated with x_{js} for $j = 1,2$ and for all s and t. Then, when a differencing operator is passed through the model,

$$\Delta y_t = \gamma_1 \Delta x_{1t} + \gamma_2 \Delta x_{2t} + \Delta \varepsilon_t \tag{29}$$

The error term in the transformed model is serially correlated with a moving average process, but the transformed regressors are still uncorrelated with the transformed errors, so the estimates remain consistent. All that is lost in this case is efficiency.

Now consider what happens with *different* degrees of differencing. Suppose the model in (28) still applies and that y and x_1 are both I(1) and cointegrated, while x_2 is I(0). The strategy described earlier applied to this example amounts to estimating the model

$$\Delta y_t = \gamma_1 \Delta x_{1t} + \gamma_2 x_{2t} + v_t \tag{30}$$

where v_t is just shorthand for the implied error term. By comparing (29) and (30), we see that $v_t = \Delta\varepsilon_t - \gamma_2 x_{2t-1}$. If there is any serial correlation in x_2, the error term in this case will be correlated with one of the regressors. The usual estimation procedure (least squares in this simple illustration) will be inconsistent.

Now, in the converse of the initial assumption, suppose the model does not apply in the levels of the variables but it does apply after the variables have been differenced to stationarity (perhaps with different degrees of differencing in the variables). In this case, the model in the differences will be consistently and efficiently estimated by maximum likelihood. The latent variable will be stationary because it is a linear combination of stationary variables, and it will be a good predictor of the normalizing indicator variable in its differenced form because the assumptions of the model are satisfied. The strategy in two of these studies is to integrate the latent variable from the differences model to become the predictor of the levels form of the normalizing indicator. Now, the latter variable has a unit root (that is why it was differenced) and the integrated latent variable will have a unit root, but there is nothing to connect these two unit roots – the two variables will not be cointegrated. Again, we have the unconvincing setting of an index that has no long-run relationship with the indicator variable that it is supposed to predict.

Giles and Tedds form the predictor by applying the coefficients estimated on the differences to the variables in the original levels. In the special case where the variables are all differenced to the same degree, this method is equivalent to integrating the latent variable from the differences model. In general, then, this method exhibits the problem described in the previous paragraph, that the predictor is not cointegrated with its target. Nor will the problem be ameliorated by different degrees of differencing. Viewed from the perspective of creating the predictor in levels from the estimates on the differences, additional unit roots are introduced when the individual variables are integrated separately and to different degrees. Again, there can be no cointegration between the predictor and its target unless the levels variables are cointegrated at the outset. Additionally, in Giles and Tedds' approach, there is a contradiction between the assumption we noted earlier that the model holds in the original levels of the variables – which implies cointegration – and the apparent need to difference the variables to avoid finding spurious relationships.

The strategy of differencing to stationarity before fitting the MIMIC model pays lip service to the issues of unit roots and cointegration, but lacks any clear purpose. To the extent that the strategy is designed to avoid spurious regressions, that objective

would be better served by less reliance on goodness-of-fit criteria (which all three studies report with gusto) and more attention to the logic of the relationships in the model. In any case, the purpose of fitting the MIMIC model is not to obtain coefficient estimates with standard asymptotic properties, nor to investigate whether significant structural relationships exist, but to condense the information contained in the indicator and causal variables into a time-series index that tracks the unobserved underground economy. That is a prediction question, and needs to be addressed by a strategy for making good predictions.

Is the mimic model appropriate?

The MIMIC model has its origins in the factor analysis of psychometrics, where the correlations of observable variables are explained by common factors or unobservable latent variables. Whether or not a statistical model is suited to a particular application is to some extent a question of judgment, but there are extensions of the original psychometric factor model where the MIMIC structure seems natural. Suppose the indicator variables are scores on various tests of ability, perhaps differentiated by subject matter such as written and verbal language and mathematics. The unobserved factor influencing all of these outcomes might be called "intelligence". In recognition of its hypothetical origins, it might be agreed to measure intelligence on a scale that for convenience is set to average 100 across the population, with a standard deviation of 15. The causal factors for intelligence will depend on the psychological theory, but they might include various parental and environmental characteristics, such as father's education and mother's nutrition during pregnancy.

This psychometric application to measuring intelligence seems far removed from estimating the underground economy in a MIMIC model. For one thing, the underground economy is not a latent or hypothetical quantity like intelligence; it is all too real, just difficult to measure because the agents who participate in it have every incentive to hide their actions. Unlike the psychometric example where the units of measurement can be resolved by convention, the concept and measurement of income in the underground economy are the same as in the observed economy. Once its scope and units are defined, the level of underground income *is* some number, calculated on a well-defined system of measurement. It cannot be open to the researcher to slide or stretch this calculation to fit whatever scale is found to be convenient. On that ground alone, the MIMIC model seems unsuited to the purpose of measuring the underground economy.

A MIMIC model relates multiple indicators y (p x 1) to multiple causes x (q x 1) through a single latent variable η (scalar). As observed by Jöreskog and Goldberger (1975), two broad implications for the observed variables follow from the assumption of a MIMIC structure. The first is that, apart from scale and some independent measurement errors, the indictors $y_1,\ldots,y_p$ are supposed to be alternative measurements of the *same* thing, namely the unobserved quantity η. The

second says that, given the causes $x_1,\ldots,x_q$ and the latent variable η, the indicators $y_1,\ldots,y_p$ are mutually *uncorrelated*. Neither of these properties is convincing in these applications to measure the underground economy.

On the first property, none of the earlier applications makes the argument that the indicator variables in their study are just noisy measurements of the underground economy, up to a scale factor. Indeed, to do so would be ludicrous because of the nature of the variables concerned. The pair of indicators in each case consists of observed GDP in some measure, and currency holdings by the public in some measure. In Giles and Tedds, the indicators are $1n(Y/P)$ and C while the index is $100U/Y$ in percentage units; in Bajada and Schneider, the indicators are $\Delta 1n(Y/(P \text{ x } N))$ and $\Delta 1n(C/(P \text{ x } N))$ while the index is $100\Delta 1n(U/Y)$ in percentage units; in Dell'Anno and Schneider, the indicators are $\Delta 1n(Y/P)$ and $\Delta 1n(C)$ while the index is $\Delta U/P$ in units of 10 billion euros. In no case in these free-form interpretations is the index even specified to be in the *same dimension* as the indicator variables, so it is impossible to sustain the idea that the indicators are just scaled and noisy measurements of the latent variable. Even if that problem were somehow fixed, it would still beggar belief to suppose that some function of observed income is an observation *of* the underground economy, just missing an adjustment for scale and clouded by errors of measurement. The same disbelief applies in parallel with the other indicator variable, which is some function of currency holdings. It makes no sense to suppose that some transformed versions of observed GDP and currency holdings are measurements of the *same* unobserved entity, whether or not that entity is called the underground economy. This foolishness is compounded in the examples of Giles and Tedds and Dell'Anno and Schneider by the use of one indicator in *real* income and the other in *nominal* currency.

The second property mentioned by Jöreskog and Goldberger indicates that the dependence structure of a MIMIC model is tightly specified. While the model is usually written in terms of covariances and linear relationships, much clearer statements can be made under the additional assumption that the variables in the model are jointly normally distributed (which assumption is implicit in estimation of the model by maximum likelihood). In particular, the correlation structure in a MIMIC model requires that (i) the indicators y are conditionally independent of the causes x, given the latent variable η, and (ii) the indicators $y_1,\ldots,y_p$ are mutually independent, given the latent variable η. Expressed less formally, these implications say that all of the connections that the indicator variables have with the causal variables, and with each other, are carried through the latent variable.

Both of these implications are unacceptable in the applications being considered here. The first suggests that observed GDP and currency holdings are related to the various causal factors in the model – tax rates, unemployment rates, government expenditures, etc. – only through the size of the underground economy. Such a proposition is inconsistent with every known macroeconomic theory of income determination. The second proposition is equally implausible because it says that

currency holdings are unrelated to observed income, once account is taken of the underground economy. If nothing else, that arrangement contradicts the currency demand model used in each of these studies to derive a benchmark value for calibrating the index from the MIMIC model.

CONCLUSIONS

We have explored the use of MIMIC models to estimate the level of underground economic activity. The three applied studies by Giles and Tedds (2002), Bajada and Schneider (2005) and Dell'Anno and Schneider (2003) have been found to be very different, despite their claims to a common parentage. Whether the MIMIC model is related to the simultaneous equations model of the econometrics textbook or the factor analysis of its psychometric origins, it is unconvincing as a framework for measuring the underground economy. The treatment of unit roots and differencing that makes this a dynamic MIMIC model is also misguided.

The literature applying this model to the underground economy abounds with alarming Procrustean tendencies. Various kinds of sliding and scaling of the results are carried out in the name of "benchmarking", although these operations are not always clearly documented. The data are typically transformed in ways that are not only undeclared but have the unfortunate effect of making the results of the study sensitive to the units in which the variables are measured. The complexity of the estimation procedure, together with its deficient documentation, leave the reader unaware of how the results have been stretched or shortened to fit the bed of prior belief.

The three applications were chosen because the datasets are available to enable replication of the calculations. No other approach would have revealed so clearly what was done to the data to obtain their estimates of underground incomes. There are many other results in circulation for various countries, for which the data cannot be identified and which are given no more documentation than "own calculations by the MIMIC method". Readers are advised to adjust their evaluation of these estimates accordingly.

Postscript – a response to the critique

The original manuscript for this paper, dated 2005, had a wide circulation. Given the delay in the paper going to publication, it seems only fair to disclose that, in the meantime, a contrary view has been put. Roberto Dell'Anno and Friedrich Schneider wrote a lengthy and detailed response to the criticisms in this paper. The bottom line of Dell'Anno and Schneider (2006) is: "According with us, the MIMIC model is still one of the best approaches to this purpose." This is not the place for me to engage in any rejoinder – readers who have followed so far will be well equipped to make their own judgements on the merits of the competing analyses and arguments.

REFERENCES

Aigner, D. J., Schneider, F., & Ghosh, D. (1988). Me and my shadow: Estimating the size of the US hidden economy from time series data. In W. A. Barnett, E. R. Berndt, & H. White (Eds), *Dynamic Econometric Modelling* (pp.224-243). Cambridge: Cambridge University Press.

Anderson, R. G., Greene, W. H., McCullough, B. D., & Vinod, H. D. (2008). The role of data/code archives in the future of economic research. *Journal of Economic Methodology, 15*(1), 99-119.

Bajada, C. (1999). Estimates of the underground economy in Australia. *Economic Record, 75*(231), 369-384.

Bajada, C., & Schneider, F. (2005). The shadow economies of the Asia-Pacific. *Pacific Economic Review, 10*(3), 379-401.

Breusch, T. (2005a). The Canadian underground economy: An examination of Giles and Tedds. *Canadian Tax Journal, 53*(2), 367-391.

Breusch, T. (2005b). Australia's cash economy: Are the estimates credible? *Economic Record, 81*(255), 394-403.

Dell'Anno, R. (2003). *Estimating the* S*hadow Economy in Italy: A Structural Equation Approach*. Working Paper 2003-07, Institut for Økonomi, Aarhus University. Retrieved from ftp://ftp.econ.au.dk/afn/wp/03/wp03_07.pdf.

Dell'Anno, R., & Schneider, F. (2003). The shadow economy of Italy and other OECD countries: What do we know? *Journal of Public Finance and Public Choice, 21*(2-3), 97-120.

Dell'Anno, R., & Schneider, F. (2006). *Estimating the* U*nderground Economy by using MIMIC* M*odels: A* R*esponse to T. Breusch's Critique*. Working Paper No. 067, Department of Economics, Johannes Kepler University of Linz. Retrieved from https://ideas.repec.org/p/jku/econwp/2006_07.html.

Frey, B. S., & Weck-Hannemann, H. (1984). The hidden economy as an "unobservable" variable. *European Economic Review, 26*, 33-53.

Giles, D. E. A. (1999). Modelling the hidden economy and the tax-gap in New Zealand. *Empirical Economics, 24*, 621-640.

Giles, D. E. A., & Tedds, L. M. (2002). *Taxes and the Canadian Underground Economy*. Toronto: Canadian Tax Foundation.

Goldberger, A. S. (1972). *Structural Equation Methods in the Social Sciences*. Amsterdam: North-Holland.

Hamilton, J. D., Zha, T., & Waggoner, D. F. (2007). Normalization in econometrics. *Econometric Reviews, 26*(2-4), 221-252.

Helberger, C., & Knepel, H. (1988). How big is the shadow economy? A re-analysis of the unobserved-variable approach of B.S. Frey and H. Weck-Hannemann. *European Economic Review, 32*, 965-976.

Hill, R. (2002). The underground economy in Canada: Boom or bust? *Canadian Tax Journal, 50*(5), 1641-1654.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, *70*(351), 631-639.

McCullough, B. D., McGeary, K. A., & Harrison, T. D. (2006). Lessons from the JMCB Archive. *Journal of Money, Credit and Banking*, *38*(4), 1093-1107.

Pagan, A. (1979). Some consequences of viewing LIML as an iterated Aitkin estimator. *Economics Letters*, 3, 369-372.

Schneider, F., & Enste, D. (2000). *Shadow Economies around the World: Size, Causes, and Consequences?* IMF Working Paper No.26, International Monetary Fund, Washington, DC.

Smith, R. S. (2002). The underground economy: Guidance for policy makers? *Canadian Tax Journal*, *50*(5), 1655-1661.

Tedds, L. M., & Giles, D. E. A. (2005). Response to Breusch's critique. *Canadian Tax Journal*, *53*(2), 392-395.

Zellner, A. (1970). Estimation of regression relationships containing unobservable variables. *International Economic Review*, *11*, 441-454.

APPENDIX– FURTHER PROBLEMS IN THE THREE APPLICATIONS

Negative variances

A difficulty that arises in all three of the studies described is that the estimates are *inadmissible*, in the sense that one of the variance estimates is negative. This outcome is obtained whether the model is estimated by the LIML/GLS procedure or by the packaged solution in LISREL®. In both Giles and Tedds (2002) and Dell'Anno and Schneider (2003), the problem parameter is ψ , which represents the variance of the latent variable. The LISREL® output file in these cases includes the prominent message "WARNING: PSI is not positive definite". It seems that the MIMIC model is *not* a good description of the data in either of these applications, despite the many measures of goodness of fit and the extensive diagnostic testing reported with the estimation results.

In Bajada and Schneider (2005), the offending parameter is θ_1, the variance of the measurement error on the first indicator variable. Again, the solution for a variance is negative, so the estimated MIMIC model is inadmissible, despite being an apparently good fit. In this case, LISREL® does not signal the problem quite so clearly, since no warning message is given. The problem is further obscured by the poor choice of units of measurement for the indicator variables. Both indicator variables in this study are quarterly *proportional* growth rates of macroeconomic variables (real per-capita income and currency holdings). These are quite small numbers, with at least one, and often two or more, leading zeros after the decimal point. The *variances* of such small numbers will be an order of magnitude smaller, because of the squaring operation in forming a variance. More than that, the parameter is the variance of the *observation error* in the variable, which will be that much smaller again. Therefore, these parameters have values that will not be readable in an output field that provides for a moderate but fixed number of decimal places, and they will be completely invisible in the default-width field of two fixed decimal places printed by LISREL®. The answer a researcher will see for each variance estimate in this case is zero. The only signal that something is wrong with the estimate of θ_1 is the negative t-ratio given for this parameter.

Other problems

Giles and Tedds (2002)

We have already noted that the (undocumented) use of standardized variables in Giles and Tedds' estimation model, together with the original variables in the prediction formula, makes the whole procedure sensitive to the units of measurement. As a complication of this sensitivity, their estimate for Canada has nothing to do with most of the causal factors in their model. As shown in Breusch (2005a), their index is almost entirely a rescaling of the variable *SELF*, which is an economy-wide aggregate measured in nominal Canadian dollars. Thus, the major part of the astounding growth they report in the underground economy over 20 years

is due to inflation in the price level, while a lesser part is due to expansion of the real size of the Canadian economy, and even less to the composition of the real economy. Nothing of their estimate can be associated with the more plausible factors that they list among their causal variables, such as the number of self-employed persons relative to the rest of the labour force, or the various tax rates. Their estimated growth rate is not even approximately a measure of the underground economy in Canada.

Also in Breusch (2005a), it is shown that the key parameters are unidentified in the currency demand model used by Giles and Tedds to calibrate the series, so the overall level of the series is not really an estimate at all, but rather a numerical accident. Vastly different "estimates" can be obtained by innocently tweaking some features of the method that should be irrelevant, such as the starting values for the nonlinear algorithm or the software package used for estimation. As with the growth rate, the level of their reported series has nothing to do with measuring the underground economy.

One additional problem in Giles and Tedds – which, in the context of the other problems, is of interest only to researchers seeking to replicate their results – occurs where a variable is not actually differenced as stated. The unemployment rate variable *M* is described as I(1), and it is reported that all integrated variables are differenced to stationarity. In contradiction to this statement, their MIMIC estimates and subsequent calculations can be replicated only if *M* is not differenced.

Bajada and Schneider (2005)

The vague language and skimpy reporting of the procedures in this paper frequently make it difficult to tell *what* is being calculated. There are also confusing lapses in accuracy. For instance, the quantities plotted in their Figure 4 are called "growth rates", and the vertical axis is labelled "%", although both of these attributes are likely to be wrong. My replication of their calculations suggests that the values plotted at an annual frequency are not annual rates of growth, as a reader might expect, but instead *quarterly* growth rates that have been averaged over the four quarters of the Australian financial year. The interpretation of the latent variable from the fitted MIMIC model as a *percentage* growth rate seems unwarranted, too, since all of the variables in the model are proportional, not percentage growth rates. Taken together, these corrections suggest that the numbers in Figure 4 should probably be multiplied by 400.

Another confusion revealed by replication of the results is a reversed set of labels in the legend of Figure 4, so what is called the "Currency-demand" line is actually the "MIMIC" result, and vice versa.

The outcomes of the calibration and integration operations are only partially revealed in the paper. In particular, the interim inference about the growth rates in Figure 4 covers only the period from 1980 to 2003, while the final inference about the levels in Table 3 is restricted to an even smaller range from 1993 onwards. There

are apparent errors even in this subset of the results, where the growth rates and the levels are mutually inconsistent. To sidestep the problem of the reversed labels in the legend of Figure 4, we can consider only periods in which the currency and MIMIC methods agree on the direction of change. However, there are instances such as the period 1993-2000, in which all of the growth rates are said to be positive, yet in Table 3 for this period there are *falls* in the levels by both methods.

Calibration in this paper is done from a slightly modified form of Bajada's (1999) currency demand model. The difference here is that the excess sensitivity measures of taxes and welfare benefits are expressed in real per-capita terms rather than percentages of GDP. Breusch (2005b) shows that the original Bajada method is highly sensitive to the units of measurement. In particular, changing the measurement of tax payments from a percentage to a proportion of GDP produces a very different inference about the underground economy (in fact, the estimates become negative!). Exactly the same objection applies in this case, where the substantive results will change when some other units of measurement are used. Replication shows that the results of the paper require the excess sensitivity variables, tax and welfare benefits, to be measured in single dollars per capita, with a 2001-02 price base. Any other scale will give a different outcome. As an example, if the variables are measured in units of thousands of dollars per capita, the results become nonsensical: the "underground economy index" of Figure 2 plummets over time until it is approximately –0.5 by the end of the period.

A second problem with the Bajada method is that the value of income velocity – to which the estimates of underground incomes are directly proportional – is set many times too high. An assumption is made that the income velocity of currency in the underground economy is equal to the ratio of income to currency in the observed economy. While this may have some superficial appeal, it ignores the very small part that currency represents in the money supply of the observed economy (currency is well under 10 per cent of M3 in Australia). Hence, the work that currency does in the generation of observed incomes is vastly overstated by this assumption. Setting the ratios of income to currency in the two sectors to be equal then transmits this exaggerated role of currency to the estimates of underground incomes. Much of the literature using currency modelling to estimate the underground economy makes a similar-looking assumption, but in these cases it is equality across sectors in the income velocity of *total money supply*. While there is some variation in this literature because of different definitions of money, the values of velocity are from one-fifth to one-fifteenth of the value assumed by Bajada. The estimates of incomes in the underground economy in Bajada and Schneider can be reduced in the same proportion.

There is an interesting claim in the paper that finding "very similar results" between currency and MIMIC models somehow validates both forms of modelling (pp.395-396). Given the two-stage processes of calibration and anchoring, as described above, it is clear that their MIMIC results have been tied directly to those of the

currency model. Both the level and the rate of growth of the underground economy in the MIMIC results are fixed to the currency model. Then, with the very small estimated rates of growth, the estimates of the levels in either case hardly move from their benchmark value, so it is no surprise that the two sets of results are similar for long periods, because the results called "MIMIC" are almost entirely drawn from the currency model. Furthermore, the similarity or otherwise of the results from the two models is hard to judge when we are shown the final outcome for only eleven of the thirty-seven years of data that are available.

Dell'Anno and Schneider (2003)

There are some small errors and inconsistencies in this paper, which become apparent on replication of the results. In particular, if the variables for tax burden, real government consumption and the rate of self-employment are percentages, as defined in the text of that paper, they should be similarly described in Appendix 1. The variables are then to be multiplied by 100. If these variables are indeed percentages, then the published coefficients indicate that the other causal variables in the preferred model are also in percentage form. The income indicator variable also needs to be multiplied by 100 to make it a percentage, but scaling of the currency variable is uncertain because there are insufficient decimal places in the published coefficient to see anything but leading zeros! Most likely, this variable is a ratio not a percentage.

The variable described in the text as "real government consumption (in percent of GDP)" is in fact G/Y, and thus the ratio of the two nominal variables. It is not G/P, as reported in Appendix 1, nor is it a more complicated variable involving multiple price indices, as might be inferred from the description in the text. The data period for estimation is unstated in the paper, but the results are most closely replicated by using 1960s1 to 2000s2 (although effectively the data begin in 1962s2 due to the creation of lags and missing observations in the currency variable).

An external estimate is used to anchor the series, so that the growth rates from the MIMIC model are converted into a time series of the level of the underground economy as a percentage of recorded GDP. The overall level of the final product of Dell'Anno and Schneider is due entirely to this external estimate, since only the variations up and down from the anchor point come from the MIMIC model. The anchor value of 19.7 per cent in 1978s2 is obtained as the simple average of five other estimates by various methods (one of which is itself the average of two others). Most of these prior estimates come from an unpublished working paper by Schneider and Enste (2000), in which they are documented as "own calculations".

A Semantic Network Analysis of Laundering Drug Money

Martin Neumann[56] *& Nicholas Sartor*[57]

Abstract

This article presents a case study of a money-laundering process. A database of police interrogations for a number of interrelated cases shows the enormous complexity of this process, exceeding the capacities of manual reconstruction. For this reason, semantic networks were reconstructed from the textual data, using the natural language processing techniques of artificial intelligence. These enabled the semantic field of this particular case to be dissected. The results reveal highly professional worldwide financial transactions. Criminal activity benefited from the infrastructure of offshore centres of the legal financial economy and permeated legal business, and the borders between legal and illegal activities became blurred. In fact, the money-laundering activity was only uncovered after the network broke down. Before the group had become known following an outbreak of internal conflict, the concealment of illegal sources of money had not been detected by law enforcement agencies. A case study does not allow for generalization. In particular, this case is not representative because the actors had access to significant resources beyond the reach of petty criminals. However, the findings from this case suggest that, in principle, professional money launderers are able to evade money-laundering regulations.

Keywords: Money laundering, layering illegal assets, text-mining, semantic networks

INTRODUCTION

Money laundering is often perceived, by both scientists and the general public, as a threat to legal society (Quirk, 1997; Steinko, 2012). In globalized times, it is assumed that criminals take advantage of the extended global financial economy. The threat scenario assumes that the intrusion of illegal money will undermine the integrity and stability of the financial system. This claim is built on the fact that legal companies are an essential part of the social and economic order, and that the social order of society is called into question by corrupt companies carrying out money-laundering activities. The transnational character of professional money laundering arguably undermines legal society. Since law enforcement is based on the authority of individual states, transnational activities can easily escape the limits of law enforcement agencies; therefore, the state is no longer the frame within which to establish social order (Steinko, 2012).

However, since money laundering is obviously undertaken in the shadows, empirical data remain sparse. Public opinion is based on speculation. Moreover, the international anti-money-laundering regime is only weakly based on scientific and empirical foundations and has to rely to some degree on *ad hoc* assumptions

[56] Corresponding author, University of Koblenz. Email: maneumann@uni-koblenz.de.
[57] University of Koblenz. Email: nicholassartor@uni-koblenz.de.

(Gilmour, 2014; Levi & Reuter, 2006; Van Duyne & Levi, 2005). A financial action task force (FATF) was established in 1989 to provide policy recommendations to fight money laundering and terrorist financing (Turner, 2015). However, the effectiveness of such measures is rarely evaluated (Levi & Reuter, 2006). In fact, statistical inspection of a Spanish example only weakly supported the threat scenario (Steinko, 2012): while this study found a few instances of transnational activities, the great majority were neither transnational nor highly professional. This result is all the more surprising since Spain is a major transit country for illicit drugs imported from Africa and destined for the European drug market. This calls into question the validity of the threat scenario.

The unclear data motivate a closer inspection of the process of money laundering. Statistical figures do not reveal details of how the process of money laundering impacts on society. For this reason, a case study is examined here, which will contribute to knowledge of the processes and mechanisms involved in money laundering. Since this is a case study, it is not representative and its findings do not allow for generalization. However, it enables in-depth insights into how money laundering is actually undertaken and how it impacts on society. The data are taken from a number of police interrogations in related criminal investigations centred on a group of drug dealers. The investigations were not undertaken to investigate money laundering, but were initiated after an outbreak of violence within the criminal group, which has been investigated using classical qualitative methods, namely a grounded theory approach (Neumann & Lotzmann, forthcoming). However, the interrogations also provide rich insights into "everyday" processes in the criminal group. While the data shed little light on the production and distribution of drugs, they reveal many details about the laundering of drug money. The capital stock was equivalent to several hundred million Euros, yet the money-laundering activities were only uncovered after the network had collapsed into violent internal conflicts. It was only as a result of investigations initiated to address the violence that the police gained access to certain information. For instance, the shock caused by the violence, which included numerous murders, may have motivated witnesses to cooperate with the police and to testify about the illegal background of certain economic activities.

However, the data consist of several hundred pages of documentation from the police interrogations. Pieces of information about the money laundering can be found scattered in various interrogations throughout the documents. It became clear on first inspection of the texts that the money laundering consisted of hundreds of activities and financial transactions in which many people all over the world had been involved to a greater or lesser degree. These data could not be handled manually with an interpretative approach. Thus, the grounded theory that we had successfully applied to a previous analysis of the escalation of violence was inappropriate for this research question. For this reason, we applied the methodology of text-mining and semantic network analysis, which has been facilitated by recent developments in information technology. Text-mining

software scans a document for relevant words and text phrases, and semantic network analysis constructs relationships between these elements (Diesner, Frantz, & Carley, 2005). In contrast to social networks, which are relations between people, semantic networks are relations between any entities that are meaningful with regard to a particular research question. This web of relations can be denoted as the semantic field of the case in question. Thus, we applied a methodology of semi-automatic, computer-guided information extraction to cope with the complexity of the data.

The paper will proceed as follows. First, the approach will be placed in theoretical context. Next, the methodological approach will be explained in some detail to show how textual data in the police interrogations were transformed into semantic networks. There then follows a detailed presentation of the results, ending with some concluding remarks.

Theoretical approach

Semantic networks and actor network theory (ANT): Dissecting the field of money laundering

While semantic networks can be analysed with measures and visualizations taken from social network analysis, there is nevertheless a difference between "semantic networks" and classical "social network analysis" (SNA). In SNA, a network is generated by "physical" (or measurable) contacts between people; for instance, if one person makes a telephone call to another person, this is proof that the two people really are in contact with each other. A semantic network is based on measures of the proximity of terms in a text document. We use the concept of co-occurrence, which assumes that two terms are meaningfully related if they appear close together. For instance, in the sentence "John is tall", the words "John" and "tall" are rather close. In a semantic network, a relationship between the semantic concepts "John" and "tall" would be constructed. However, if two terms appear together in a piece of text, it does not prove that the terms are in "contact" with each other or have something in common. It may be just by chance that in a text document certain terms appear quite near to each other; the relationship is not based on physical activity, such as making a phone call, or on a personal statement that person A says he is a friend of person B. For instance, in the sentence "I think of John and the moon is red", the words "John" and "moon" are close together; nevertheless, they are not related. There is thus a danger of false positives. Later in this paper, we give a more technical explanation of the research process, showing how we attempted to minimize the likelihood of such errors.

On the other hand, the technique has advantages over classical SNA. Classical SNA consists simply of the persons involved in a network. The "ontology", that is, the description of the domain, is restricted to persons, and thus restricts the information that can be extracted from SNA. The situation is different in *semantic* networks. Concepts may be anything for which a word exists. The "network of concepts" is

not restricted to persons. For this reason, much more information can be extracted than in classical SNA. This relates to features of actor network theory (ANT). While differences undoubtedly exist between semantic networks and classical approaches to ANT, there is one important commonality: ANT treats actors and non-actors, and material and non-material concepts, equally. This is also true of semantic network analysis. There are therefore parallels in the objectives and results of ANT and semantic network analysis.

ANT cannot be reduced to a single coherent theory, but is rather a bundle of various studies originating from science and technology research. However, the core objective of ANT has been described as follows:

> You do not go about doing your business in a total vacuum but rather under the influence of a wide range of surrounding factors. The act you are carrying out and all of these influencing factors should be considered together. This is exactly what the term actor network accomplishes. An actor network, then, is the act linked together with all of its influencing factors (which again are linked), producing a network (Hanset & Monteiro, 1998, Ch.6).

In this way, ANT attempts to reconstruct how human and non-human elements interact. This has been called a material semiotic approach (Latour, 2005) which explores the relational ties between concepts that constitute a certain field. These concepts may be a multitude of different entities. For instance, Isaac Newton did not invent the theory of gravitation on his own. He relied on observational data from astronomers (who are human actors), as well as on support from the Royal Society (which is a socially-constructed value; money, for example, can be described as a semiotic entity since its value as an exchange medium depends on social agreement) and his room in Trinity College (which is material). All these together were constitutive of the scientific innovation ascribed to Newton (Bardini, 2000). Thus, all these elements constitute an actor network: "the actor is the network of heterogeneous relations that [are] able to redefine and transform what it is made of" (Callon, 1987, p.93). While it is certainly true that important differences exist, in this regard semantic networks share a central objective with ANT, namely to dissect relationships between concepts from different ontological domains. The concepts in semantic networks are not only actors (as in social network analysis), but also include, for instance, tasks, resources and human actors. Construction of the network enables us to investigate how these heterogeneous elements are related in the constitution of a certain field, in our case the field of money laundering. The basis of the network is the text. Thus, the network describes semiotic relations: constructs related by their meaning.

Social and human capital

Classical SNA has been widely applied in criminological research (Baker & Faulkner, 1993; Duijn, Kashirin, & Slot, 2014; Klerks, 2001; Krebs, 2002; Sparrow, 1991). It is particularly appropriate for analysing the *social* capital of actors (Bourdieu, 1984; Wassermann & Faust, 1994) because analytic concepts such as degree centrality or betweenness centrality enable one to identify actors holding strategic positions in a network (Duijn et al., 2014). Degree centrality reveals central hubs in networks by identifying those actors who have more contacts with more actors. By contrast, betweenness centrality measures broker positions, by identifying actors who hold positions between different cliques. These actors can bridge structural holes or, in other words, enable contacts between otherwise disconnected parts of the overall network (Wassermann & Faust, 1994).

The innovation of semantic networks is to enable the extraction of information about linkages between human actors and non-human, material and non-material aspects related to the field of interest. This has advantages for the purpose of examining criminal activity. For instance, semantic networks provide an opportunity for the analysis of *human* capital. The concept of human capital originates from economic theory and denotes the specific knowledge, skills and competencies of employees involved in a production process. The same concept can be applied to criminal activities (Sparrow, 1991). Depending on the degree of professionalization, money laundering may involve specialized competencies in financial transactions. Knowledge and access to specific resources can be identified by analysing relationships between human actors and non-material elements such as tasks or resources.

The field: Money laundering

In our case, the field of investigation is money laundering. For criminals to enjoy or to re-invest the profits of their business, it is important that illegal money is transformed into legally usable wealth. In recent decades, the global dimension of money laundering has attracted increasing attention. Although the source of illegal wealth may be any illegal activity, the most important is drug trafficking (Harnischmacher, 2009). Money laundering is the process of legalizing illegal assets (Schneider, Dreer, & Riegler, 2006). According to Schneider et al. (2006), four goals can be distinguished. The first goal is to hinder confiscation by the public authorities, while nevertheless retaining control over the money, which is the second goal. For this purpose, the money is imported into the legal financial economy, which is the third goal. The final goal is protection from criminal prosecution. Typically, it is assumed that money laundering consists of three phases: placement, layering and integration. This three-phase model was first developed by the US customs authorities and has since been adopted worldwide by, for example, the United Nations Office on Drugs and Crime (Levi & Reuter, 2006).

Placement is the process through which illegal assets are introduced into the financial system in the first instance. Often it involves breaking down huge amounts of money into small portions. Placement can take various forms, such as investment in amusement halls, insurance companies or real-estate projects. Other options are to bring money physically outside the country of origin to areas with less well-protected financial markets – so-called offshore financial centres. These are often countries such as Netherland Antilles or Bermuda, but Jersey and some Swiss cantons may also be considered. In the course of globalization, these offshore centres gained in importance, not only to the legal financial economy but also to illegal financial activities (Coats & Rafferty, 2007). However, in this phase a direct link between the illegal source and the first placement in the legal market still exists; thus, the origin of the asset can still be retraced. For this reason, the second stage, layering, is essential.

Layering is the process of obscuring the sources of the money, making the investor anonymous in order to foil criminal prosecution. This can be done in various ways, and often involves multiple financial transactions. If these are undertaken across borders, it becomes increasingly unlikely that the source will remain visible (Harnischmacher, 2009).

Integration is the final step that reallocates the money back into the hands of the investor, with the origin of the profit now hidden from public view. This ensures that the goal of retaining control over the asset is achieved. Like layering, this process is also open to the creativity of the actors. As examples, the literature mentions real-estate trade, the acquisition of companies (often companies with huge cash flows, such as restaurants, amusement halls, etc.), and fictitious financial transactions that, again, may result in investment in legal economic activities (Harnischmacher, 2009; Levi & Reuter, 2006).

In achieving these goals, financial intermediaries such as banks, investment funds and insurance companies, and also "underground banks", play a crucial role (Levi & Reuter, 2006). For this reason, the financial economy is assumed to be vulnerable to being undermined in illegal ways. As early as 1989, the FATF was established to protect global financial systems from money laundering, financing terrorism and other financial crimes (Levi & Reuter, 2006; Turner, 2015). Forty recommendations, including reporting suspicious transactions, were originally formulated to increase financial transparency. Since the FATF was first set up, its mandate has been constantly expanded. However, the FATF is an instrument of soft law: it is an agreement between states, but does not have the authority of international treaties that can provide concrete prescriptions (Brumer, 2010). Rather, it is a committee of experts that monitors states and provides recommendations. Thus, the case study presented here can be regarded as a test case for the efficiency of the soft law prevention of financial crime.

DATA AND METHODOLOGY

The database consisted of police interrogations in a number of interrelated cases. Documentation on the interrogations was delivered by the police to the researchers after being anonymised using software tools. However, the data are not publicly available, so the protection of privacy was ensured. Basically, the procedure consists of two parts: text mining and the extraction of semantic networks (Diesner & Carley, 2010; Diesner, Carley, & Tambayong, 2012; Diesner et al., 2005; McCallum, 2005). This procedure reveals relational information from textual data, and is carried out in several consecutive steps. First, data pre-processing is required, followed by node identification and edge identification. The tools used were AutoMap for text mining and ORA for network construction; both were developed at the Center for Computational Analysis of Social and Organizational Systems at Carnegie Mellon University. They were chosen because they are connected by a common interface, which facilitates the integration of text mining and network analysis, and because ORA is a network analysis tool that provides more analytical power than pure text-mining tools.

However, methodologically, police interrogations pose a challenge for analysis because of the demands of extracting a network from a huge corpus of unstructured textual data in natural language. It turned out that a great deal of manual work and control was necessary, since information extraction tools are highly language sensitive. Ready-made dictionaries that facilitate text mining exist in more prominent applications of text-mining technologies, such as sentiment analysis in Twitter feeds (Voinea & Schatten, 2015). However, they do not exist for very specialized domains such as money laundering. From the perspective of natural language processing (NLP), it appears that "the state of the art methods in natural language processing are still not robust enough to work well in unrestricted text domains to generate accurate semantic representations of texts" (Aggarwal & Zhai, 2012, p.3). This statement is corroborated by the fact that, in the analysis that was performed, manual and automatic data analysis were tightly interwoven (Sartor, 2015).

Text Mining

Text mining aims to extract information from texts. Technically, pre-processing of the data was undertaken first, followed by identification of the basic forms of concepts and classification according to a meta-ontology. All these steps involved the mutual interplay of manual and computational analysis.

First, pre-processing consists of error cleaning and the removal of non-content-bearing concepts (words that are only grammatically relevant). Words eliminated by a so-called "stop word filter" included words like "the". Punctuation marks were also eliminated, and words were transformed to lower case. This was undertaken automatically. Metadata were also eliminated, including, for example, pagination and data that had been included by the police, such as the names of police

investigations. Errors in the translation and compilation of the data also had to be resolved. The original document was a scanned PDF file that had to be transformed with OCR tools before it could be imported into the analysis software. Errors occurred during this process. For instance, the number "5" was sometimes identified as the character "S" by the OCR tool. These steps in the pre-processing were undertaken manually.

Next, to identify basic units of information, words that contained essentially the same information were reduced to a common form. In particular, different spellings of the names of persons, locations and organizations were reduced to a single form. For example, the abbreviation "Ltd." had sometimes been used to denote a company, and sometimes not; and in the witness statements, an individual such as "John Smith" (a fictitious example) had sometimes been called "Mr. Smith" or just "John". Moreover, words that together formed a semantic unit (technically n-grams) were combined using an underscore character, for example the transformation of "several thousand dollars" to "several_thousand_dollars". The text-mining software created a concept list of more than 13,000 concepts that had been detected automatically, and included suggestions about the form into which the concepts might be transformed (denoted as "concept from" and "concept to"). Concepts forming the basic unit of information ("concept to") were checked and revised manually to maximize accuracy. Often, this involved going back to the original document to inspect the context.[58]

The next central step was to classify these words according to a meta-ontology taken from the literature (Diesner & Carley, 2005). This meta-ontology defines the classes of objects with which the world can be described. The world consists of the following classes:

Agents: This means simply persons.

Organizations: These are mostly companies.

Resources: In principle, this is something from which somebody can gain something. It turned out to be rather difficult to classify terms consistently as resources. With respect to money laundering, resources are mostly flows of money, but they are also real estate ownership, etc.

Tasks: These are efforts that somebody has to make, that take place over at least a certain length of time. Paying rent for a flat would be a task, whereas buying a company would be classified as an event.

Groups: These are people teaming up together. In our data, this turned out to be unimportant.

Events: An event is everything that is (or can be) associated with a date.

[58] Whereas in larger bodies of text this problem can partly be resolved by machine learning technologies, in our case the body of text was too small and heterogeneous for such technologies, even though it was too big and complex for qualitative research.

Locations: These are places, but are of different sizes. Locations range from whole countries, such as Switzerland or Curacao, to street numbers.

However, only those words were selected that belonged or might be relevant to money laundering. As the interviews were originally conducted to investigate the outbreak of violence in the criminal group, many words could obviously be found that described acts of violence. Since the focus of this text analysis was money laundering, these words were ignored in the analysis. Thus, words such as cash, company, contract, etc. were included, but *not* words such as murder, intimidation, etc. In fact, the meta-ontology also contains the class "unknown". All words that appeared to be irrelevant to the task of analysing money laundering were assigned to this category and deleted from the analysis. A few concepts in the "unknown" class were retained, as it was unclear to the authors whether or not they were relevant. Deleted words were replaced by "xxx". This had the advantage that the distance between relevant concepts was preserved, a feature that became important in the development of the semantic network. Applying the delete list reduced the number of concepts from 13,511 to a final number of 4,123. Note that the assignment of concepts to the meta-ontology was undertaken by a mutual interplay of automatic and manual work. Initial suggestions were made by the analysis software. However, the error rate was high and the detection rate was small – most concepts were classified as unknown; therefore, the classification was revised manually. This was done by one author and cross-checked by the other author to minimize misleading classifications.

In the network diagrams, the different classes of the ontology appear in different colours (see Figures 3 to 7). Agents are red, resources are turquoise, tasks are blue, locations are dark red, and organizations are green. First, we show the count of concepts preserved for the subsequent network analysis (see Table 1).

Table 1. List of concepts in the data

Classes	Nodes
Agent	357
Event	892
Knowledge	16
Location	353
Organization	472
Resource	1,202
Role	190
Task	380
Unknown	72
Attribute	173
Group	16

Semantic Network Analysis

Before undertaking a network analysis, the relationships between the words needed to be identified. Since the analysis was based on police interrogations, the network revealed what the police suspected to be involved in money laundering. Network identification is based on the concept of co-occurrence, when two terms appear quite near to each other in the text. What "quite near" means can be manually adjusted, so that, for example, no more than five or no more than ten words must lie between two terms. After initial experiments, a window size of five was applied. At this stage, it became important that deleted terms had been replaced by "xxx" in the text-mining phase of the research, so that the distance between the terms had been preserved. The assumption of a measure of proximity based on co-occurrence is that two words are related in terms of semantic content if they appear close together. However, to reduce the likelihood of false positives, only those relationships that appeared together a number of times were used for the network analysis. The assumption was that it was unlikely that a chance co-occurrence would happen frequently. After initial experiments with different frequencies, the frequency was set to at least 10. Applying this filter also reduced the number of isolated nodes. Co-occurrence is the basis for edge identification from textual data. Once edges are identified, networks of edges can be constructed. These are based on the meta-ontology, insofar as a network can be constructed, for example that relates actors and organizations. These networks can then be analysed by measures of classical network theory such as degree centrality or the construction of ego networks.

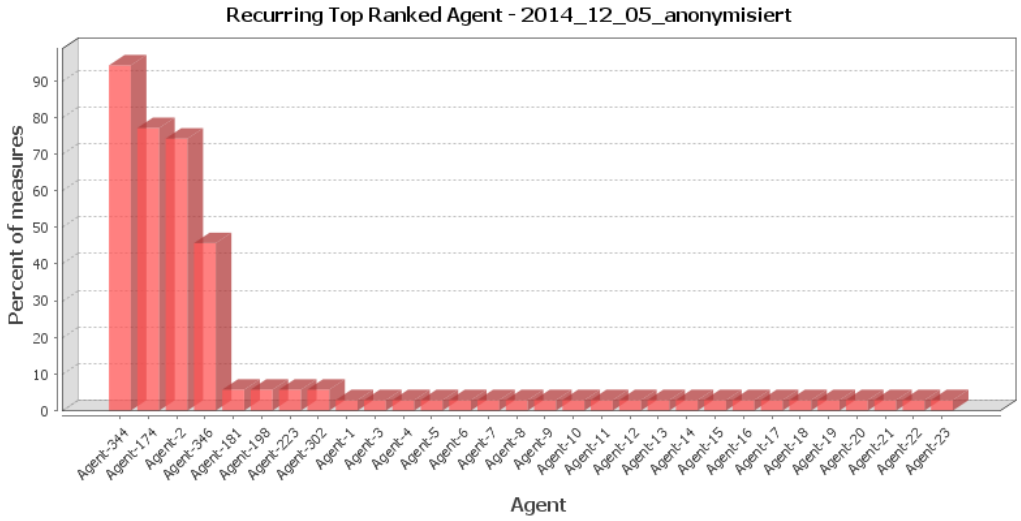
RESULTS

In this section, networks of interest with regard to money laundering are documented and briefly explained. It should be noted that the networks cannot be regarded as proof of money-laundering activities, particularly not in juridical terms. However, they reveal patterns that provide hints about how the activity was organized. This limitation is also partly due to the fact that the data were anonymous. Police investigators with access to the non-anonymous version were able to connect the broad patterns with their additional experience based on detailed knowledge of a particular case.

First, as the most basic analysis, the central actors and organizations were identified. These guided the following analysis, insofar as the central actors and organizations appeared frequently in the later networks. In the next steps, the organizational structure was revealed in more detail. This provided the basis for tracing the money flow throughout the organizational structure in a number of further analyses. Details emerged of the roles of specific persons and locations: for example, the network of relations of the concept of money laundering was merged with the network of relations of one of the central actors, and the offshore location Curacao was

examined in close detail. Finally, the semantic field of the associated concepts was investigated by a clique analysis.

Figure 1. Ranking of agents according to degree centrality



First, the central actors were identified based on their degree centrality, that is, on having the highest number of edges. Figure 1 reveals that a small number of actors were central hubs in the network. The five central actors were agents 344, 174, 2, 346 and 181. However, Figure 1 shows immediately that agent 181 was far less central than the top four. The same was done for the organizations found in the data, as shown in Figure 2. The ranking of the organizations also reveals that certain organizations were the most important. The three most important were organizations 31, 447 and, unsurprisingly, the bank. However, the decline in degree centrality is smoother than in the case of the actors. This is a first indicator that the central actors were involved in multiple organizations.

Figure 2. Ranking of organizations according to degree centrality

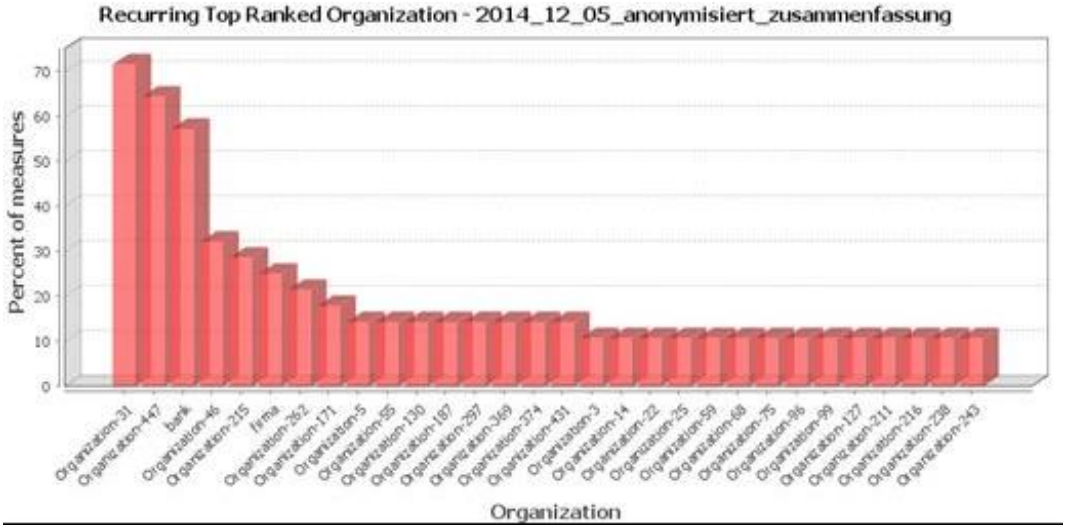
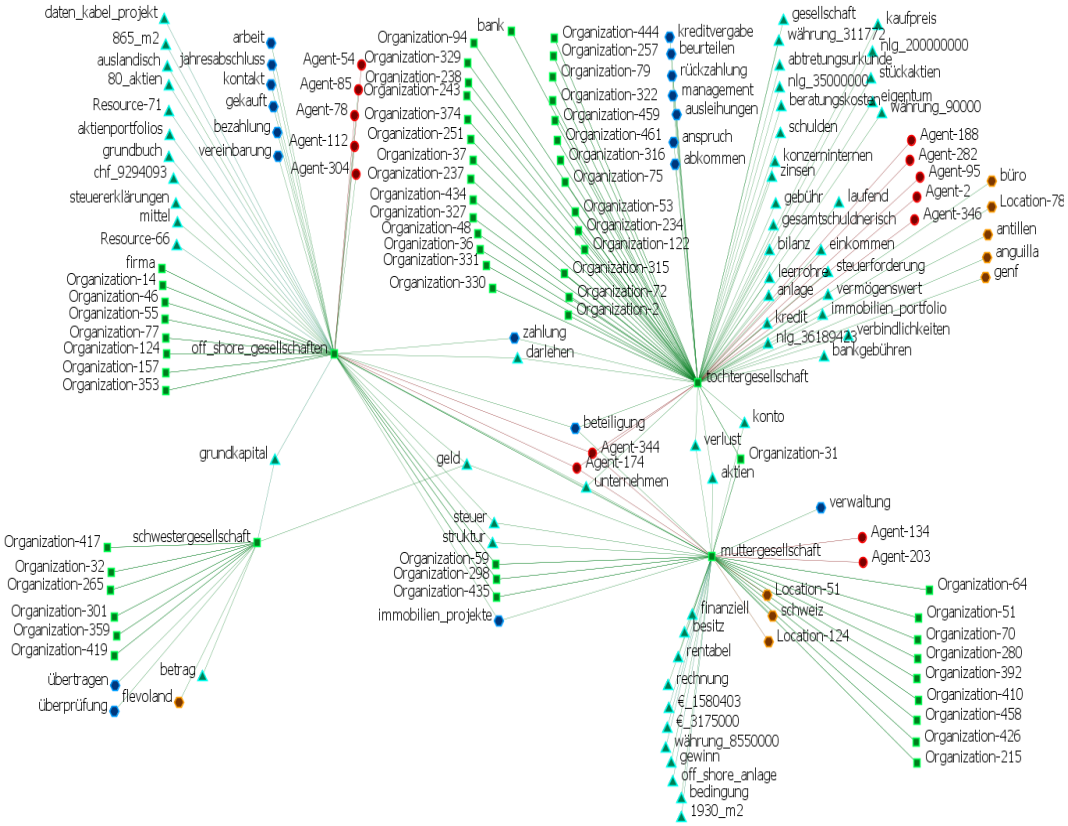


Figure 3 provides details of how the "system" of organizations was structured. It shows how the organizations were related to one another, as well as to the actors, resources and (some) tasks.

Figure 3. Structure of the network of companies

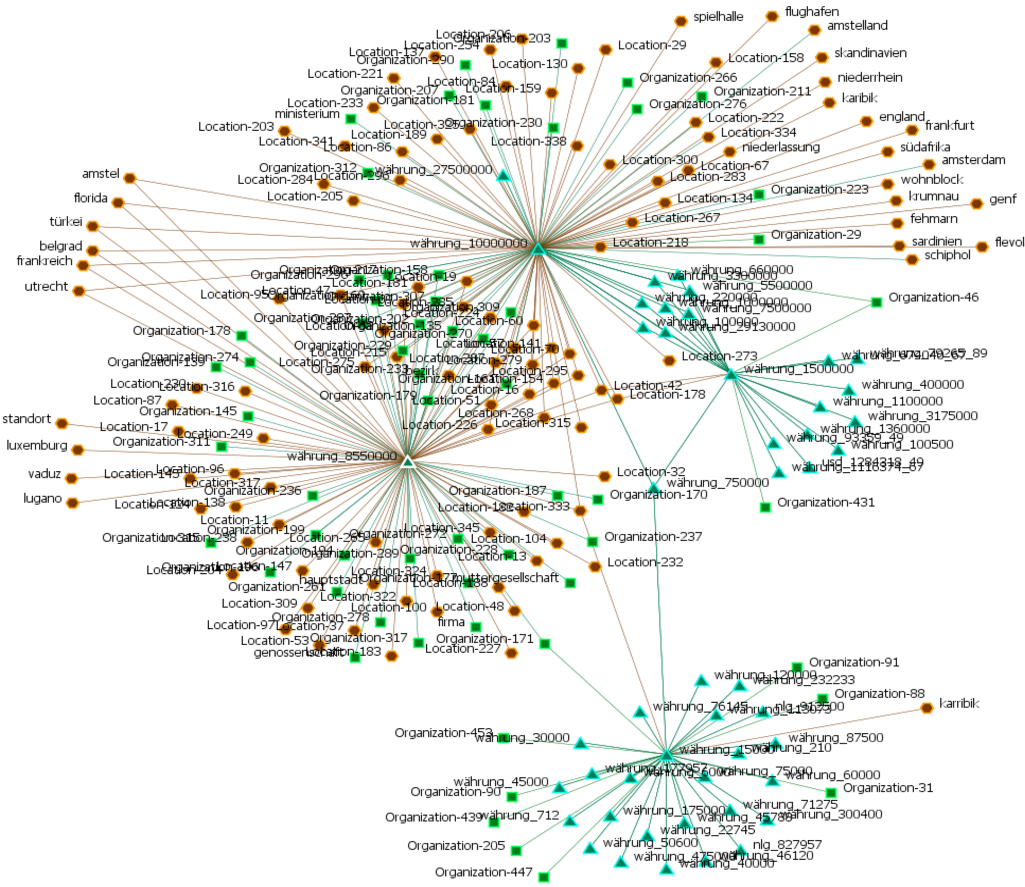


First of all, it becomes apparent from Figure 3 that the organizational structure consisted of a complex web of interrelated companies. The central concepts of the organizations are parent companies ("muttergesellschaft"), subsidiary companies ("tochtergesellschaft"), affiliate companies ("schwestergesellschaft") and offshore companies. The central parent organization had control of various organizations, such as organizations 64 to 215 which can be seen in the lower right corner of the figure. The parent is directly linked to offshore companies, indicating the importance of offshore centres. In between the concepts of the parent and subsidiary companies, we find organization 31. This organization is identified as the most central according to its degree centrality. We see that it plays a specific role by relating the parent company to its daughter companies. The companies are bound together by means of shares ("aktien") and losses ("verlust"). These concepts can be found in the figure just next to organization 31. Thus, Figure 3 explains how organization 31 gained centrality in its role in the network of companies.

The affiliate companies ("schwestergesellschaft") are connected to the parent companies via the concept of money ("geld"), which can be found in the middle of the figure between these two concepts. The affiliate companies are also related to

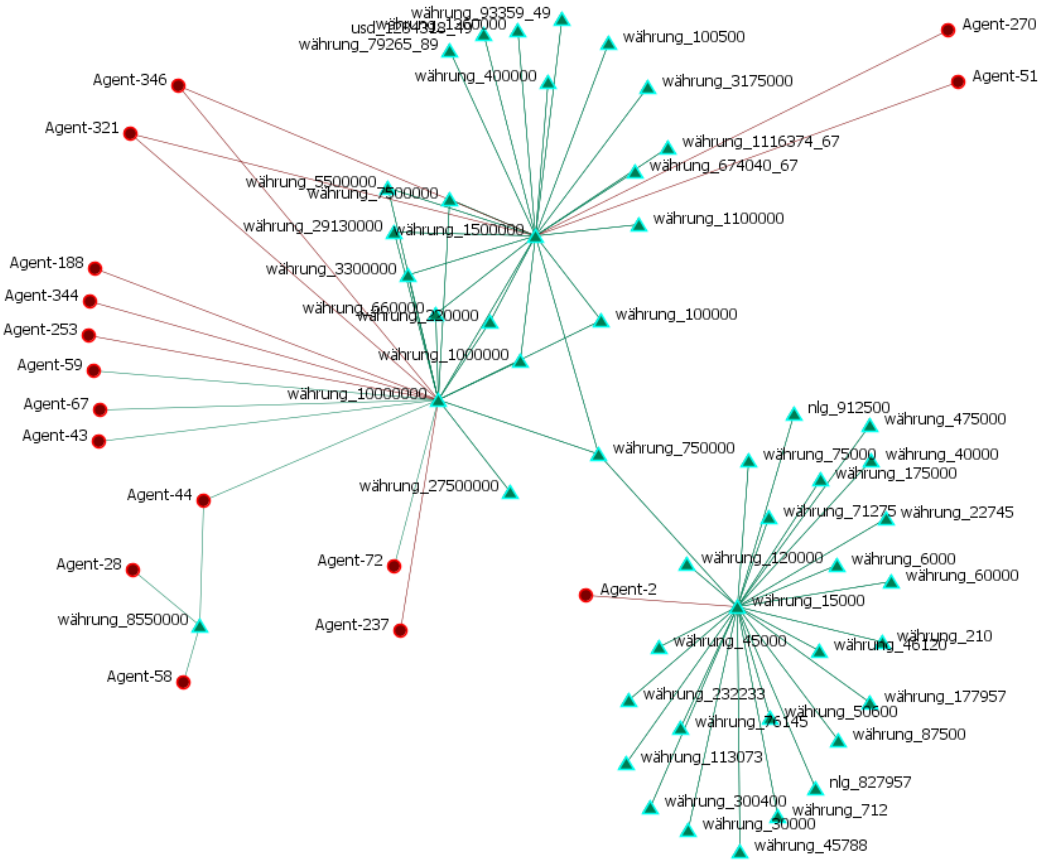
the offshore companies by the holding of capital (“grundkapital”). Thus, capital seems to be the central binding force in the company network, with shares and losses as ties between the parent company and the subsidiary companies, and equity holding (“beteiligung”) connecting offshore centres, the parent organization and subsidiary companies. Similarly, the subsidiary companies are connected to the offshore companies via payment (the blue node of the task “zahlung”) of loans (the resource “darlehen”). Note, furthermore, that the concept of real-estate project (“immobilien_projekte”) is connected to the parent company and to the offshore companies, as can be seen by the blue dot at the bottom in the middle, indicating a transfer of tasks of projects within the alliance of firms. Such a transfer of tasks to offshore companies hints at the purpose of concealment. This is typical of tax evasion, and also of money laundering, being the process of layering as described in the money-laundering cycle. Moreover, real-estate business is a well-known method of integration in the process of money laundering.

Figure 4. Money flow



Reconstructing the network of the company structure indicates a high degree of professionalism. In the centre are agents 344 and 174, who are the two most important actors according to their degree centrality. They are directly linked to the concepts of parent company, subsidiary company and offshore company. In this way they have key access to the resources of the company network.

Figure 5. Management of concealment



Concealment is investigated in greater detail in Figure 4. This shows a network of resources, locations and organizations. The first thing to note in Figure 4 is a cloud of densely connected organizations and locations on the left in the middle of the figure. The cloud is connected to specific amounts of money ("währung 8,550,000" on the left and "währung 10,000,000" at the top). The dense cloud between these two hubs reveals the complex flow of these specific amounts of money through various organizations in different locations. This suggests that the objective was to make the source untraceable. In fact, it was not detected by previous financial investigations. The specific amounts of money act as hubs for locations, organizations, and even other resources, which enables inferences to be drawn about the money flow. The amount of "8,550,000" and the amount of "10,000,000" have a number of ties to companies and, to a limited degree, this also holds for the amount of "15,000". Moreover, hubs for certain amounts of money ("währung 1,500,000" and "währung 15,000") have direct ties to other concrete amounts of money. This indicates a differentiation between different individual amounts of money. Some seem to play a central role, giving rise to other financial transactions. Again, these specific amounts of money have been transferred through the organizations with which they are related. This indicates an attempt to obscure the source of the money, which suggests that these financial transactions are part of the layering in the money laundering activity.

Figure 5 investigates the details of this activity, and is a network of resources and agents. The resources are currencies ("währung"), or money. This network shows how specific financial resources are tied to specific agents. In particular, we wish to draw attention to the relationship between agent 2 and the amount of 15,000. As we know from Figure 1, agent 2 is one of the central actors in the network. From reading the text in a qualitative way, it became apparent that the amount of 15,000 was related to a number of interrogations of a person who was paid 15,000 per month in return for personal consultancy services in relation to financial issues for one of the criminals. Neither the concrete business field of the company in which this individual worked, nor his official profession, were uncovered in the data. However, presumably he worked as something like a tax consultant. In the interviews, it was clear that he did not regard himself as a criminal. In fact, his activities did not violate any law. Several times he insisted that this extra money – paid in addition to the usual payments from his employer – was not something unusual. Thus, this person was a perfectly law-abiding businessman. Nevertheless, the huge cluster of other amounts of money related to the hub of "currency 15,000" (i.e. his monthly payment) indicates that he nevertheless played a central role in the money laundering. These other values refer to the financial transactions that he organized. Apparently, the service provided in return for the monthly payments was the organization of the money flow in the laundering activities. Note that, once the transactions had been undertaken by a legal businessman working for a reputable and legal company, the money became integrated into the legal economy, which is a prime goal of money laundering.

Here, we find a case in which illegal money permeated the legal economy through a legal businessman, and thereby corrupted the legal economy, at least to a certain degree. While the financial activities were not illegal, it remains questionable whether the individual should have accepted a significant payment in return for this professional service without further consideration. It could be said that he was paid for "not asking questions". Obviously, the financial transactions are suspicious by FATF standards (Turner, 2015). These standards have been in existence since 1989 and their aim is to combat money laundering. However, they have been implemented as soft law, and do not have the same legal authority as international treaties. While it was a clear violation of the informal conventions of FATF that agent 2 did not report these activities, it was not a violation of any legal obligation. Moreover, the case of agent 2 reveals the central role of *human capital* in the professional performance of the financial transactions. Specialized competencies are needed that enable the execution of specialized tasks. These activities require skills that can only be obtained through a specialized education.

Apart from agents 344 and 346, most agents in Figure 5 are not central hubs in Figure 1; thus, they play a different role from that of agent 2. This sheds light on the exploitation of *social* capital. These agents are also related to specific amounts of money: a number of agents are related to "währung 10,000,000" and two agents (346 and 321) are related to "währung 1,500,000". While agent 344 was one of the

key agents controlling the overall company structure, the other agents might have been some kind of straw men, holding, for instance, positions in letterbox companies in offshore locations. This indicates a considerable degree of specialization of tasks in performing the overall activity of money laundering, involving various forms of human and social capital. For instance, qualitative analysis revealed that the straw men often had intimate personal relationships with the central agents, being their brothers or girlfriends. These were actors who were highly trusted by the central agents because of their personal ties. Nevertheless, they remained distant from the central hubs in the professional money-laundering network, with only a few ties to the professional network. This feature made these people appear innocent, hiding their involvement in the activities of the core network that carried on professional money laundering in the background.

Figure 6. Ego network of money laundering and agent 344

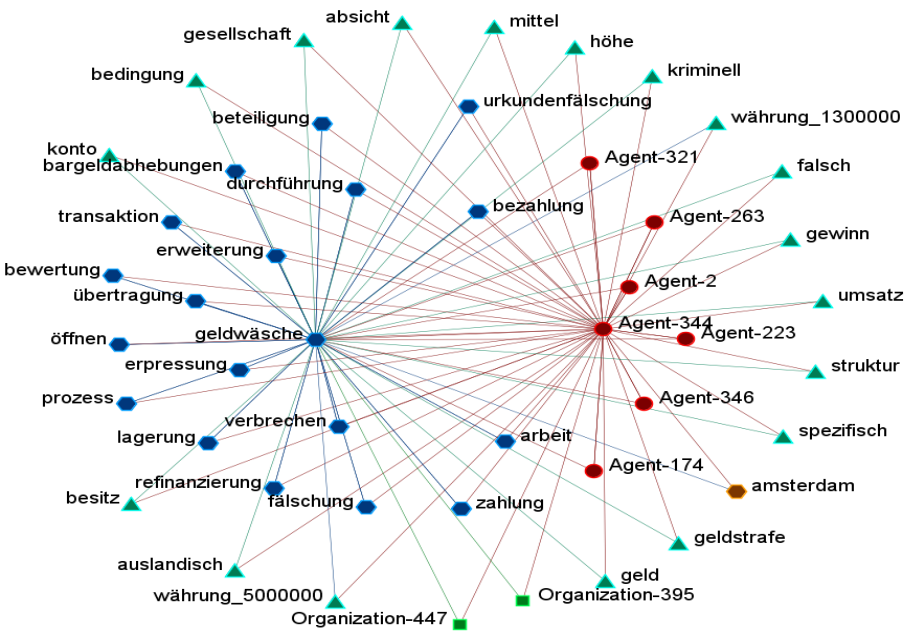
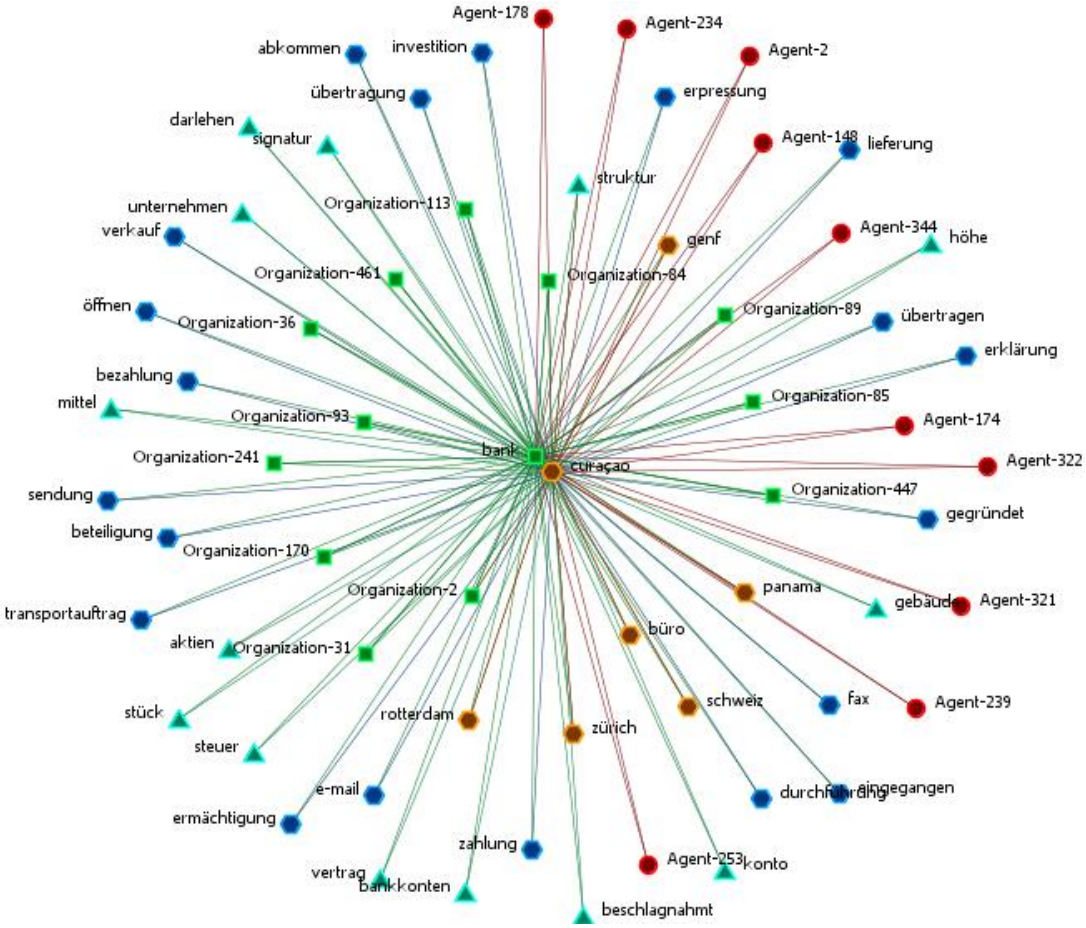


Figure 6 is a conjoint representation of the ego networks of the concepts agent 344 and money laundering ("geldwäsche"). As we know, agent 344 was one of the central actors controlling the company structure. While the ego network of an agent corresponds to the analytical concept of classical SNA, the ego network of a semantic concept such as money laundering can be regarded as a disentangling of the semantic field relating to a particular social activity. This is an attempt to examine issues relating to the topic of money laundering ("geldwäsche") in greater detail. In terms of ANT, it can be described as the actor network of money laundering in which human and non-human elements interact. Disentangling this semantic field reveals that the ego networks relating to the task of money laundering

refer mainly to the resources and organizations involved in this activity. Only one location appears in this network. Compared with the huge number of companies in the overall company structure (as shown in Figure 3), it is rather surprising that only two organizations, organizations 447 and 395, have direct ties to both agent 344 and the concept of money laundering. This suggests that the company structure displayed in Figure 3 also had a legal business aspect into which money laundering activities had been inserted. However, as we know from Figure 2, organization 447 is one of the central organizations in the overall network, and the number of other people in the network is relatively small. All have direct ties to agent 344, as well as to the concept of money laundering; they are clustered around these two concepts. It is striking that the resources are less often the concrete currency values seen in the earlier networks. Only one currency value appears in this network, "währung 1,300,000". Instead, the resources describe the attributes and properties needed to undertake the activity of money laundering, such as criminal ("kriminell"), false ("falsch"), volume of sales ("umsatz"), and asset ("gewinn"). Similarly, the tasks include activities that are typical of classical economic activities such as cash withdrawal ("bargeldabhebung", a task that is also typical of illegal financial activities), assessment ("bewertung"), transaction ("transaktion"), equity holding ("beteiligung") and refinancing ("refinanzierung"). However, the tasks also include crime ("verbrechen"), blackmail ("erpressung") and forgery ("fälschung"), which are clearly concepts from the criminal field. This feature makes money laundering a double-faced activity.

Figure 7 shows the ego network of Curacao, one of the most prominent offshore centres. The concept is shown in relation to actors, locations, tasks and resources. Remarkably, the concept "bank", which, according to Figure 4, is one of the most important organizations in the overall network, is intimately related to "Curacao". The two other most central organizations detected in Figure 2, namely organizations 31 and 447, also appear in relation to the concept "Curacao". It is also interesting to look at other locations, including Panama, another offshore location, and also European centres such as Rotterdam and the Swiss cities of Zurich ("zürich") and Geneva ("genf"). This seems to confirm the traditional assumption that the money will ultimately end up in a Swiss bank account. Again, the central agents 2 and 344 can be found. Note that agent 2 is the professional businessman discussed in Figure 5. This is a hint that Curacao played a central role in the professional management of the financial transactions. However, the other actors do not belong to the central hub shown in Figure 1. The fact that these actors appear to be linked to the concept of an offshore centre such as Curacao is a further indicator of the existence of straw men, as already suggested by Figure 5. The tasks confirm the assumption that activities related to the concept of Curacao are related to money laundering. Activities related to services in the financial economy are even more dominant than in Figure 6. The only criminal task is blackmail ("erpressung"). The other tasks include, for instance, payment ("zahlung"), sale ("verkauf"), venture ("unternehmen") and investment ("investition").

Figure 7. Ego network of Curacao



Finally, a clique analysis was undertaken to identify patterns of related concepts. Using the Newman algorithm, cliques with at least 15 members were extracted. Cliques are patterns of concepts that have a high number of internal, but only a small number of external links. Thus, they are cohesive subgroups of the overall network with a certain degree of closure. It can therefore be assumed that they belong together (Wassermann & Faust, 1994). In contrast to classical SNA, these cliques involve all the kinds of concepts that can be found in the text. So a clique can be regarded as a semantic field associated with the task of money laundering: 31 cliques with at least 15 elements were identified. For instance, clique 1 consists of 17 companies and a specific amount of money "Waehrung 1,600,000", which is an indicator of a flow of this amount of money through these organizations. It has to be noted that the cliques are not fully closed – ties exist between cliques, and the cliques are not mutually exclusive – so concepts may be part of several cliques. Organizations 57, 84, 256 and 442 appear, for instance, in 30 of the 31 cliques identified. The same holds for "£_9000" and "€_1728.06". It is striking that the latter is a very specific amount of money that is part of many cliques. Again, this suggests that this sum had been the subject of multiple transactions. These are indicators of an intention to conceal the source of the money. Finally, since the concepts of "buying" and "selling" are obviously central to financial transactions, the relationships of these concepts will be described: "buying" is related in the data

to 22 places, 36 persons and 76 organizations, and "selling" is related to 71 places, 58 persons and 101 organizations. Interestingly, "buying" only has links to 38 different sums of money, and "selling" refers to only 61 different sums of money. Thus, both concepts are related to far fewer concrete values than organizations. This does not prove anything: it might have multiple causes. However, it might well be the case, if more organizations than concrete values are at play, that one and the same amount of money was transferred through various organizations. This is a classical pattern of layering in the process of money laundering in order to conceal the source of the money.

CONCLUSION

In this investigation, a semantic network was constructed from unstructured textual data. A central difference between classical SNA and semantic networks is that concepts are not restricted to actors. The ontology is far bigger and enables more detailed and differentiated insights into the field of investigation. One can simply examine what turns out to be relevant. In the police interrogations, all words that might be related to money laundering were selected and classified according to a meta-ontology for network construction. Investigation of the relationships between semantic concepts grants an insight into how entities with different ontological statuses are related in the semantic field of money laundering. This approach is similar to ANT, insofar as it describes networks between heterogeneous concepts. Human and non-human, as well as material and non-material concepts are treated equally as they constitute concepts in the field of money laundering.

As a result, a complex structure of companies involved in highly professional financial transactions was uncovered. The relationships between actors and resources enabled us to describe the human capital invested in these activities. The example of the professional businessman (agent 2) shows how legal business became corrupted by criminal enterprise. With regard to the money-laundering cycle of placement, layering and integration, the focus of the data was on layering, namely obscuring the source of assets. This is indicated by the huge number of complex financial transactions. The organization of a complex company structure, as outlined in Figure 3, indicates a high degree of professionalism. This is further substantiated by the involvement of professional consultants in financial issues, as revealed in Figure 5. The data show that, for this purpose, the same offshore centres were used as in the official financial economy. This shows that the FATF regulations have not been effective in preventing money laundering.

Indeed, this corresponds closely with the threat scenario outlined at the beginning of this paper: the threat of professional and transnational criminal activities undermining the legal economy and society. Since this is a case study, it does not allow for any generalizations. It is likely that most money-laundering activities do not follow this scheme, simply because it requires more skill and resources than could be supplied by petty criminals. Professionalism is costly. On a small scale, a

business such as selling used cars might be a more realistic option. From the statistics, non-professional and rather local activities are more likely to prevail, as the data from Spain indicate. However, public opinion is particularly occupied by the impact on society. The case investigated here is indeed a case of the illegal economy undermining legal business. This brings into question whether statistical and causal relevance should be differentiated: a small number of extreme cases might have a bigger impact than many cases of petty crime. The case demonstrates that, in principle, professional money laundering is a challenge for regulations aimed at preventing it.

Acknowledgement. This research received funding from the European Union's Seventh Framework Program (FP7/2007-2013) under grant agreement no. 315874, GLODERS Project.

REFERENCES

Aggarwal, C., & Zhai, C. X., Eds (2012). *Mining Text Data.* New York: Springer.

Baker, W., & Faulkner, R. (1993). The social organization of conspiracy: Illegal networks in the heavy electrical equipment industry. *American Sociological Review, 58,* 837-860.

Bardini, T. (2000). *Bootstrapping: Douglas Engelbart, Coevolution, and the Origin of Personal Computing.* Stanford, CA: Stanford University Press.

Bourdieu, P. (1984). *Distinction. A Critique of the Judgement of Taste.* Cambridge, MA: Harvard University Press.

Brumer, C. (2010). Why soft law dominates international finance – and not trade. *Journal of International Economic Law*, *13*(3), 623-643.

Callon, M. (1987). Society in the making: The study of technology as a tool for sociological analysis. In W. Bijker, T. Hughes, & T. Pinch (Eds), *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology* (pp.83-103). Cambridge, MA: MIT Press.

Coats, N., & Rafferty, M. (2007). Offshore financial centers, hot money and hedge funds: A network analysis of international capital flows. In L. Assassi, A. Nesvetailova, & D. Wigan (Eds), *Global Finance in the New Century* (pp.38-54). Basingstoke: Palgrave MacMillan.

Diesner, J., & Carley. K. (2005). Revealing social structure from texts: Meta-matrix text analysis as a novel method for network analysis. In V. Narayanan, & D. Armstrong (Eds), *Causal Mapping for Research in Information Technology* (pp.81-108). Hershey, PA: IDEA Group Publishing.

Diesner, J., & Carley, K. (2010). Extraktion relationaler Daten aus Texten. In C. Stegbauer, & R. Häussling (Eds), *Handbuch Netzwerkforschung* (pp.507-521). Wiesbaden: VS Verlag.

Diesner, J., Carley, K. M., & Tambayong, L. (2012). Extracting socio-cultural networks of the Sudan from open-source, large-scale text data. *Computational and Mathematical Organization Theory*, *18*(3), 328-339.

Diesner J., Frantz T., & Carley K. (2005). Communication networks from the Enron email corpus: "It's always about the people. Enron is no different". *Computational and Mathematical Organization Theory*, *11*(3), 201-228.

Duijn, P., Kashirin, V., & Slot, P. (2014). The relative ineffectiveness of criminal network disruption. *Nature.com Scientific Reports*, *4*, Art. No. 4238. doi:10.1038/srep04238.

Gilmour, N. (2014). Understanding money laundering: A crime script approach. *The European Review of Organized Crime*, *1*(2), 35-56.

Hanset. O., & Monteiro, E. (1998). *Understanding Information Infrastructure* (online manuscript). Retrieved from http://heim.ifi.uio.no/~oleha/Publications/bok.html.

Harnischmacher, R. (2009). Geldwaesche. *Die Kriminalpolizei: Zeitschrift der Gewerkschaft der Polizei.* Retrieved from http://www.kriminalpolizei.de/ausgaben/2009/ detailansicht-2009/artikel/geldwaesche-money-laundering-1.html.

Klerks, P. (2001). The network paradigm applied to criminal organizations. *Connections, 24*, 53-65.

Krebs, V. (2002). Mapping networks of terrorist cells. *Connections*, *24*, 43-52.

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor Network Theory.* Oxford: Oxford University Press.

Levi, M., & Reuter, P. (2006). Money laundering. *Crime and Justice*, *34*(1), 289-375.

McCallum, A. (2005). Information extraction: Distilling structured data from unstructured text. *ACM Queue, 3*(9), 48-57.

Neumann, M., & Lotzmann, U. (forthcoming). Eine Herrschaft des Terrors: Gewalteskalation in illegalen Organisationen. In A. Groenemeyer, C. Equit, & H. Schmidt (Eds), *Situationen der Gewalt.* Weinheim: Beltz.

Quirk, P. (1997). Macroeconomic implications of money laundering. *Trends in Organized Crime, 2*(3), 10-14.

Sartor, N. (2015). *Eine Netzwerkanalyse der Organisationsstrukturen im Bereich der Geldwäsche.* Master's thesis, Universität Koblenz-Landau, Koblenz, Germany.

Schneider, F., Dreer, E., & Riegler, W. (2006). *Geldwaesche: Formen, Akteure, Groessenordnungen – und warum die Politik machtlos ist.* Wiesbaden: Gabler.

Sparrow. M. (1991). The application of network analysis to criminal intelligence: A transnational organized crime threat assessment. *Social Networks*, *13*, 251-274.

Steinko, A.F. (2012). Financial channels of money laundering in Spain. *British Journal of Criminology*, *52*, 908-931.

Turner, N. (2015). The financial action task force: International regulatory convergence through soft law. *New York Law School Review*, *59*, 548-559.

Van Duyne, P., & Levi M. (2005). *Drugs and Money: Managing the Drug Trade and Crime – Money in Europe.* London: Routledge.

Voinea, C., & Schatten, M. (2015). Recovering the past: Eastern European web mining platforms for reconstructing political attitudes. *European Journal of Political Attitudes and Mentalities, 4*(1), 22-39.

Wassermann, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press.

Drivers of Suspicious Transaction Reporting Levels: Evidence from a Legal and Economic Perspective

Julia Braun[59], *Matthias Kasper*[60 61], *Alicja Majdanska*[62], *Maryte Somare*[63]

Abstract

Suspicious transaction reporting (STR) is a cornerstone of the international Anti-Money Laundering/Combatting the Financing of Terrorism (AML/CFT) framework. The evaluation of AML/CFT regimes is challenging, however, as the quality of STRs varies substantially between countries and little is known about the factors that drive STR. In combining legal and economic analyses, this article evaluates various factors that potentially explain STR levels. The analysis of the AML/CFT legislation in nine jurisdictions reveals that well-established legal and institutional structures promote the effectiveness of STR systems. In particular, the legal analysis shows that the scope of predicate offenses in national criminal law, as well as a penalty regime for non-compliance with the obligations under national AML/CFT legislation, seem to increase the quantity of STRs. Overly strict penalty regimes and insufficient training of entities with reporting obligations, on the other hand, likely stimulate over-reporting. Based on these findings, we econometrically investigate potential determinants of STR levels for 54 countries from 2006 to 2012. We find that high STR numbers indicate high levels of illegal activities such as terrorism and organized crime. Moreover, mutual evaluations of countries' AML/CFT frameworks entail a short-term increase in the number of STRs.

[59] ZEW, Centre for European Economic Research, L7, 1 68161 Mannheim, Germany, DIBT Fellow.
[60] Department of Applied Psychology: Work, Education, Economy, University of Vienna, Universitätsstraße 7, 1010 Vienna, Austria.
[61] WU Global Tax Policy Center, Vienna University of Economics and Business (WU), Welthandelsplatz 1, 1020 Vienna, Austria.
[62]Institute for Austrian and International Tax Law, Vienna University of Economics and Business (WU), Welthandelsplatz 1, 1020 Vienna, Austria.
[63]Christian Doppler Laboratory for Transparency in International Tax Law, Institute for Austrian and International Tax Law, Vienna University of Economics and Business (WU), Welthandelsplatz 1, 1020 Vienna, Austria.

The authors wish to thank Michael Lang, Jeffrey Owens, Rick McDonell, Sergej Teterukov, Selvan Lehmann, Juan Sebastian Castano Gardeazabal, Olena Dudar Mustafa Yeter, Zareh Asatrayan, Chris Heady and two anonymous referees, and participants at the Shadow Economy, Tax Evasion and Fiscal Intermediaries Conference at TARC, University of Exeter, and at the WU Conference Tax and Good Governance in Africa for helpful discussions and comments. Julia Braun and Matthias Kasper gratefully acknowledge financial support by the Austrian Science Fund (FWF) W 1235-G16. Matthias Kasper gratefully acknowledges financial support by the Austrian National Bank (OeNB) 16042. The views expressed in this paper as well as all remaining inaccuracies and errors are, of course, solely our own.

INTRODUCTION

Combining legal and econometric methods, this paper analyses suspicious transaction reporting levels. While the number of these reports differs substantially between countries, little is known about the factors that explain a given level of reporting. In identifying and analysing a broad set of legal and economic factors that potentially contribute to countries' reporting levels, this article aims to facilitate the evaluation of suspicious transaction reporting schemes.

In response to the growing threat of global terrorism, the Financial Action Task Force (FATF)[64] initiated the development of a global standard of policy and legal measures to fight money laundering, terrorist financing and the proliferation of weapons (hereinafter the FATF Standard) (FATF, 2012). FATF Recommendation 20 proposes the introduction of a reporting standard that requires national financial intelligence units (FIUs) to report suspicious transactions related to criminal activity or terrorist financing. Depending on national legislation, these suspicions usually take the form of suspicious transaction reports (STRs) or suspicious activity reports (SARs).

The efficiency of STR systems depends on two factors: the quality of reports and the administrative capacity to detect and prosecute criminal activity. Both factors, however, rely on the implementation of several legislative and bureaucratic provisions (Roule & Kinsell, 2003). Therefore, periodic evaluations of the degree to which jurisdictions are compliant with the FATF Standard are an integral part of the global anti-money laundering/combating of the financing of terrorism (hereinafter AML/CFT) framework. Mutual evaluations are conducted by financial experts and found to be objective and consistent (Chaikin, 2009). Moreover, when conducting mutual evaluations of AML/CFT regimes, the question often arises as to what an adequate level of such reporting might be. As reporting entities are held liable for the consequences of money laundering, FIUs often receive a large number of unfounded STRs (e.g. Johnston & Carrington, 2006; KPMG, 2003), which makes the FIUs' evaluation particularly important.[65] Experience of the assessors has shown that the number of STRs could range from hundreds a year to thousands a day. Against this background, it is very difficult to make any meaningful and informed judgment about whether or not a given level of reporting is adequate in a country's particular context. However, in order to facilitate the evaluation of FIUs efforts in fighting money laundering and the financing of terrorism, a better understanding of the effectiveness of STR schemes and the factors that drive reporting levels is much needed.

[64] The FATF is an inter-governmental body established in 1989 by the Ministers of its member jurisdictions. The objectives of the FATF are to set standards and promote effective implementation of legal, regulatory and operational measures for combating money laundering, terrorist financing and other related threats to the integrity of the international financial system.

[65] FIUs often forward only a small fraction of STRs filed to law enforcement agencies and sometimes several thousand STRs refer to one money laundering case (e.g. see EUROSTAT, 2013).

While it should be acknowledged that direct comparisons between different countries are not feasible due to each country's unique circumstances (legal, financial, etc.), we assume that some factors may contribute to and explain a given level of reporting. In combining legal and economic analyses, this article aims to evaluate various factors that potentially explain STR levels. Improving the effectiveness of AML/CFT regimes is crucial as global terrorism and international money laundering remain significant threats to political stability and economic prosperity in many regions of the world.

Our main findings are threefold. First, the legal analysis reveals that the scope of predicate offenses in national legislation and a dissuasive penalty regime for non-compliance with national AML/CFT regimes seems to increase the number of STRs. On the other hand, insufficient training of reporting entities, as well as overly severe penalties for non-compliance with reporting standards are likely to promote over-reporting. Second, the econometric analysis shows that illegal activities, particularly terrorism and organized crime, trigger high levels of STRs. Third, both legal and economic analyses find evidence that mutual evaluations increase the number of STRs.

In line with earlier research on the effects of FATF membership on money laundering in the banking sector (Johnson & Lim, 2003)[66], these findings suggest that the FATF Recommendations facilitate the detection of criminal activity related to terrorist financing and money laundering when implemented appropriately. Moreover, we observe that STR is particularly effective when embedded in sound legal structures with a broad scope of predicate offenses. Despite the resources required to build and maintain capacity to detect suspicious activities, the FATF Recommendations appear to play a crucial role in fighting global terrorism. However, as we find the number of STRs to be closely linked to mutual evaluations, mutual evaluation schemes should incorporate risk considerations. This would help monitoring jurisdictions that are vulnerable to money laundering and terrorist financing more effectively.

The next section describes the STR requirements as set out in the FATF Recommendations. We subsequently investigate legal and institutional factors that affect STR levels and discuss the effectiveness of different reporting standards in Section 3. In Section 4, we assess the effects of various economic and institutional factors on STR levels. Section 5 summarises and concludes.

[66] While Johnson & Lim (2003) find that countries exhibit less money laundering activity in the banking sector after joining the FATF relative to non-FATF members, they do not investigate whether the FATF aggravates the money laundering problem in non-member states when illicit activity is relocated to non-FATF countries. Also, their research design does not allow identifying the direct effects of FATF membership on money laundering activity.

International standards in suspicious transaction reporting

Suspicious transaction reporting is the backbone of preventive measures under the FATF Standard. Specifically, the FATF recommends legal provisions that require financial institutions (FIs)[67] and designated non-financial businesses and professions[68] to report certain transactions to their national financial intelligence units (FIUs). The most relevant recommendation in this respect is Recommendation 20, which reads as follows:

> *"If a financial institution suspects or has reasonable grounds to suspect that funds are the proceeds of a criminal activity, or are related to terrorist financing, it should be required, by law, to report promptly its suspicions to the financial intelligence unit (FIU)."* (FATF, 2012, p. 21)

The interpretative note to Recommendation 20 clarifies its scope (FATF, 2012, p. 80). Criminal activity under this recommendation should be understood as any

[67] The term "*Financial institutions"* under the FATF Recommendations 2012 has a broad meaning and includes both natural and legal persons who are engaged in a business encompassing, or carry out on behalf of the customer, one or more of the following activities: (i) acceptance of deposits and other repayable funds from the public; (ii) lending, (iii) financial leasing, (iv) money or value transfer services, (v) issuing and managing means of payment (e.g. credit and debit cards, cheques, traveller's cheques, money orders and bankers' drafts, electronic money), (vi) financial guarantees and commitments, (vii) trading in money market instruments (cheques, bills, certificates of deposit, derivatives etc.), foreign exchange, exchange, interest rate and index instruments, transferable securities, commodity futures trading, (viii) participation in securities issues and the provision of financial services related to such issues, (ix) individual and collective portfolio management, (x) safekeeping and administration of cash or liquid securities on behalf of other persons, (xi) otherwise investing, administering or managing funds or money on behalf of other persons, (xii) underwriting and placement of life insurance and other investment related insurance, (xiii) money and currency changing. The FATF Recommendations 2012 are available online at: http://www.fatf-gafi.org/media/fatf/documents/recommendations/pdfs/FATF_Recommendations.pdf.

[68] Under the FATF Recommendations the term *"Designated non-financial businesses and professions"* is defined with an exhaustive list of subject persons that are expected to have an increased probability to be exposed to the proceeds of crime in their regular professional activity or business: (i) casinos, (ii) real estate agents, (iii) dealers in precious metals, (iv) dealers in precious stones, (v) lawyers, notaries, other independent legal professionals and accountants – this refers to sole practitioners, partners or employed professionals within professional firms. It does not refer to 'internal' professionals that are employees of other types of businesses, nor to professionals working for government agencies, who may already be subject to AML/CFT measures, (vi) trust and company service providers refers to all persons or businesses that are not covered elsewhere under these Recommendations, and which as a business, provide any of the following services to third parties (a) acting as a formation agent of legal persons, (b) acting as (or arranging for another person to act as) a director or secretary of a company, a partner of a partnership, or a similar position in relation to other legal persons, (c) providing a registered office; business address or accommodation, correspondence or administrative address for a company, a partnership or any other legal person or arrangement, (d) acting as (or arranging for another person to act as) a trustee of an express trust or performing the equivalent function for another form of legal arrangement, (e) acting as (or arranging for another person to act as) a nominee shareholder for another person. Further, it should be noted that designated non-financial businesses and professions have been included in the FATF Recommendations in 2003. The FATF Recommendations 2003 are available online at: http://www.fatf-gafi.org/media/fatf/documents/recommendations/pdfs/FATF%20Recommendations%202003.pdf.

predicate offense for money laundering as defined by the national laws of the individual countries. The FATF provides for a minimum list of offenses which should be covered by national AML/CFT legislation and be included in the definition of the predicate crimes for money laundering or terrorist financing.[69]

However, not all offenses falling in one of the designated crime categories have to be covered by the reporting requirement. FATF Recommendations[70] defining the scope of predicate offences for money laundering and terrorist financing are closely related to the STR framework. *Recommendation 3 "Money Laundering Offence"*[71] and *Recommendation 5 "Terrorist Financing Offence"*[72] outline the minimum scope of crimes that countries should include in their national AML/CTF regimes. From the wording of Recommendation 3, it is possible to derive that national laws have to define a threshold for predicate offenses that constitute serious crimes. An exception is made for terrorist financing as the *"serious crime"* threshold is omitted. Furthermore, it is stressed that terrorist financing should be understood in a comprehensive manner. This means financing of terrorist acts does not need to be linked to a specific terrorist act or acts to be considered as such (FATF, 2012, p. 80). All suspicious transactions must be reported to the FIUs, whether they have been carried out or just attempted (FATF, 2012, p. 80).

Although the FATF Recommendations do not constitute a legally binding instrument under international law, they have been globally recognized and therefore can be considered to be soft international law (Terry, 2010). In order to comply with the FATF Standard countries are expected to implement the FATF Recommendations in their national legal systems in compliance with a regular procedure laid down in their constitutional laws. Avoiding prescriptive provisions in the FATF Recommendations helps to provide for a flexible base that can be adapted to the legislative and legal specifics of each country. Such flexibility allows for a broad spectrum of designs and interpretation of the FATF Standard in the national institutional and legal AML/CFT frameworks. The enforcement of the FATF Standard is ensured by on-site visits and off-site reviews of the documentation provided by reviewees. Mutual evaluations are conducted by

[69] A list of designated offenses under the FATF Recommendations 2012 includes the following categories of offenses: participation in an organised criminal group and racketeering, terrorism, including terrorist financing, trafficking in human beings and migrant smuggling, sexual exploitation, including sexual exploitation of children, illicit trafficking in narcotic drugs and psychotropic substances, illicit arms trafficking, illicit trafficking in stolen and other goods, corruption and bribery, fraud, counterfeiting currency, counterfeiting and piracy of products, environmental crime, murder, grievous bodily injury, kidnapping, illegal restraint and hostage-taking, robbery or theft, smuggling (including in relation to customs and excise duties and taxes), tax crimes (related to direct taxes and indirect taxes), extortion, forgery, piracy, and insider trading and market manipulation.

[70] Other recommendations also may have influence on the STR reporting frameworks, but in the authors' opinion their impact is rather indirect. Therefore, the focus of the legal analysis will be on the above-mentioned three recommendations with some references to other potential influencing factors.

[71] Old numbering: Recommendation 1 and Recommendation 2.

[72] Old numbering: Special Recommendation II.

assessors who are appointed either directly by FATF, or by connected organizations, such as the Committee of Experts on the Evaluation of Anti-Money Laundering Measures and the Financing of Terrorism (Moneyval), the Asia/Pacific Group on Money Laundering (APG) and the other FATF-like regional bodies.

Legal analysis

Introduction

Since the FATF Recommendations do not constitute a legally binding instrument under international law, jurisdictions may choose their own legal and institutional models in order to implement the FATF Standard. Given a broad variety of possible legal and institutional frameworks which may be implemented by jurisdictions, nothing else may be expected but a slighter or major deviation from the international standard.

With the aim of identifying relevant features in legal and institutional frameworks that may have affected STR levels, we focus our analysis on two contrasting groups of countries from the sample surveyed in the economic analysis. STR to GDP ratios rather than absolute figures are referred to in order to distinguish high and low-reporting countries for the purpose of reviewing legal and institutional frameworks. Particularly, we focus on the three countries with very high (Thailand, Latvia, Georgia) and four countries with the very low (Qatar, Switzerland, Germany, and Austria) STR/GDP ratios, while also accounting for practices from other interesting jurisdictions such as Mexico (medium STR/GDP levels) and Liechtenstein.[73] By reviewing the mutual evaluation reports (MERs) which are publicly available on the FATF or FATF-like bodies' websites, an attempt is made to identify changes in legal systems or institutional measures which could have affected STR levels in the period from 2006 to 2012.

Scope of the predicate offenses under domestic laws

In accordance with Recommendation 20, when deciding whether to report a suspicious transaction or activity, first of all, the entity or person has to identify whether a transaction has any potential connection with proceeds derived from committing a predicate crime. Thus, the number of reported transactions clearly depends on the scope of the definitions of money laundering and terrorist financing in national laws. It is hence plausible to expect that countries that define money laundering in a broad sense by including in its definition the whole range of designated crimes can be expected to have high STR levels. This assumption seems to be confirmed when looking at the countries with low STR/GDP levels.

As depicted in Appendix A, Qatar was the country with the lowest STR/GDP ratio in the timespan analysed. Indeed, the scope of money laundering offences was very

[73] Until 2009, the OECD considered Liechtenstein as an uncooperative tax haven. http://www.oecd.org/countries/liechtenstein/listofunco-operativetaxhavens.htm.

narrowly drafted and accordingly criticized by the assessors in the MER of 2008. At that time under Qatari AML law, money laundering offenses did not cover acts aimed at concealing or disguising the location, disposition, movement, or ownership of funds and the list was further narrowed by adopting a list of predicate crimes substantially shorter than that of FATF-designated offences. Pursuant to this list, only crimes involving drugs and dangerous psychotropic substances, forgery and counterfeiting, and imitations of notes and coins, illegal trafficking of weapons, ammunition and explosives, terrorist crimes, and extortion and looting were considered to constitute crimes predicated to money laundering (MENAFATF, 2008, p. 9). As a result of these and other deficiencies Qatar was placed in a regular follow-up process (MENAFATF, 2012, p. 1). The deficiencies related to the scope of money laundering offence were removed in 2010 (MENAFATF, 2012, p. 2). Accordingly, the number of STRs significantly increased in 2010.[74]

Switzerland, the country with the second lowest STR/GDP in our sample, in 2005 had a definition of predicate offenses with a limited scope, which did not include four categories of predicate offenses specified by the FATF (human trafficking, piracy and counterfeiting, smuggling, insider trading and market abuse) (FATF, 2005, p. 13). The first three categories were addressed at the time of the follow-up report and the last one was partly dealt with (FATF, 2009, p. 16). The analysis of Swiss laws revealed that there was no requirement to file an STR if no relationship with the customer had been established due to failed negotiations. Furthermore, Swiss law required that a person filing a STR had a well-founded suspicion – a threshold that demands more verification from the person reporting a transaction than prescribed by the FATF Recommendations.

In the case of Germany the assessors noted in the 2010 MER that the concept of *"suspicious transaction"* was defined in a very narrow way. Its scope related only to money laundering and not generally to the proceeds from criminal activity (FATF, 2010, p. 165). Moreover, this definition required a much higher certainty threshold than is intended by the wording *"suspects or has reasonable grounds to suspect"* provided in the FATF Recommendations. National law indicated that the requirement of near-certainty had to be met before filing the report. Additionally, given that the German FIU was organized in the form of a police unit and STRs were submitted to criminal investigation authorities, STRs represented criminal complaints instead of constituting a preventive measure (Cindori, 2013).

An interesting example is represented by Thailand - the country with the highest STR/GDP ratio in our sample. The high ratio of reports in Thailand is not a concomitant of the definition of money laundering, but is rather attributable to inefficiencies of its AML/CFT framework (International Monetary Fund, 2007, p. 168). In fact, Thailand was rated as being partly compliant with Recommendation

[74] This effect cannot be entirely attributed to the extension of the scope of predicate crimes, but should rather be seen as the result of comprehensive AML/CFT framework reform requested by the MENAFATF.

1 and Recommendation 2 (now Recommendation 3) in the 2007 MER, where it was asked to extend the scope of predicate offences. Thailand is rather an exception, since its high reporting levels may be explained by deficiencies of its legal and institutional frameworks that are explained in more detail in Section 3.5.

On the other hand, a broad definition of money laundering crimes can lead to more adequate STR levels An example is Latvia,[75] where the level of STRs was low due to the fact that reporting entities deliberately complied with the AML/CFT framework in a very lax way (Moneyval, 2006, p. 7). Reporting levels improved, however, after regulatory reforms were undertaken in order to address international objections and to restore trust in the financial sector.

Penalties for non-compliance with reporting obligations

Sanction regimes for non-compliance with reporting obligations are central elements of AML/CTF regimes. In line with traditional economic theory, the cross-country analysis supports the assumption that high penalties imposed on non-compliant entities can work as a dissuasive factor and increase the level of STRs (Allingham & Sandmo, 1972).

Low sanctions or no fines effectively imposed seem to affect STR levels negatively. The problem of insufficiently dissuasive sanctions is common to many countries with low reporting levels. For example, in the 2008 MER of Qatar the assessors have highlighted that the Qatari AML law did not provide for sanctions for failure to report suspicious transactions (MENAFATF, 2008, p. 138). Furthermore, no institutions were sanctioned for non-compliance with AML law (MENAFATF, 2008, p. 140). Subsequent to this critique Qatari laws were amended and penalties were increased, and additional administrative sanctions were introduced (MENAFATF, 2012, p. 16). In fact, the number of STRs reported in 2010 compared to the 2009 figures increased more than two times.

Also in the case of Georgia, STR levels increased substantially due to the stricter regulatory pressure by the National Bank of Georgia on the financial sector. In fact, the number of STRs more than doubled in 2010 compared to 2009. However, at the same time financial institutions interviewed by the assessors have revealed that the substantial increase in STRs was a signal of defensive over-reporting following inspections carried out by the Georgian National Bank and penalties imposed as a result of those inspections rather than a sign of effective identification of suspicion (Moneyval, 2012, pp. 194-195).

[75] The first MER completed in 2001 already stated that "Latvia has taken a number of significant steps to counter money laundering. They have put together a very comprehensive structure for the protection of the financial system in a regime based on suspicious and unusual transaction reporting, the latter with a range of different reporting thresholds for various institutions" (Moneyval, 2001, p. 2).

In Thailand, persons subject to reporting faced a lack of feedback or guidance on how to detect suspicious transactions or which circumstances might indicate the possibility that transactions have an underlying ML/FT purpose (International Monetary Fund, 2007, p. 168). On the other hand, high sanctions and penalties for non-compliance, which in the AML/CFT framework always means the failure to report a suspicious transaction, tended to fortify further the tendency to file STRs without a well-grounded suspicion. Given the high level of uncertainty with respect to their obligations, financial entities reported more suspicious transactions than necessary (International Monetary Fund, 2007, p. 168).

Visits of Evaluators and Regulatory Pressure

Generally, the analysis of the majority of MERs proves that on-site visits of the assessors play a significant role in triggering the reforms necessary to bring the legal system of a particular country in line with the FATF Standard and increase the number of meaningful STRs. One of the most salient examples is Latvia. Being caught in the midst of the scandals revolving around its financial sector in 2005 and subject to heightened international attention,[76] including that of AML/CFT framework assessors, Latvia had to introduce a range of legislative, institutional and administrative measures to rehabilitate the image of its financial sector. Furthermore, authorities have undertaken extensive inspections at the financial institution; as a result, at one point 13 of 23 Latvian financial institutions were under intensified supervision due to AML/CTF deficiencies. It was expected that deficiencies identified during this period were to be substantially addressed by the end of 2005. As a result of these reforms and intensified pressure on the financial sector by the national authorities, the numbers of suspicious and unusual transactions reports have increased in the period of the reforms and kept on growing several years thereafter. Figure 1 represents the dynamics of the reporting levels in Latvia between 2000 and 2010 (the data are taken from Moneyval, 2012, p. 77). The 2012 MER underlined Latvia's substantial progress in improving its AML/CFT system. However, it was still rated as *partially compliant* with a range of recommendations concerning the establishment of client risk profiles and STR (Moneyval, 2006, p. 245). Generally, in countries with low reporting levels visits of assessors led to an increase of STRs due to the reforms that had to be implemented in order to comply with the observations made in the MERs (also see Section 4).

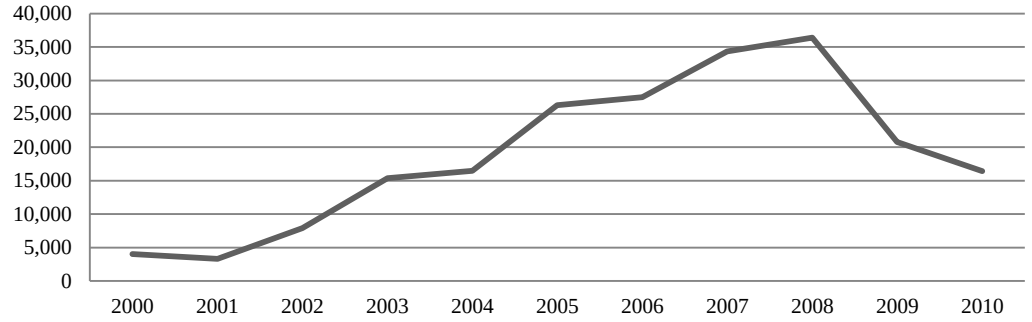
Efficiency of AML/CFT Frameworks

High levels of STRs do not automatically indicate that the AML/CFT system is entirely compliant with the FATF Standard. In 2008, Mexico's MER detected a long list of deficiencies in its AML/CFT framework, among which an

[76] The US Treasury made a proposal for classifying two Latvian banks as posing major concern for money laundering for the purposes of the US law (Moneyval, 2006, p. 7).

overwhelming quantity of low quality STRs (FATF, 2008, pp. 184-185). The latter deficiency was not only identified by the assessors, but also acknowledged by the Mexican authorities. STRs peaked in 2005, and then started to decrease due to efforts undertaken to increase the quality of the reports. It is interesting to note that although the quantity of STRs was high, Mexico did not make use of the STRs filed by some participants on the financial market, for example bonding companies, registered money transmitters and currency exchanges (FATF, 2014b, pp. 20-21).

Figure 1. Suspicious and unusual transactions reported in Latvia in 2000- 2010



The high level of reporting in Thailand could also to some extent be attributed to an inefficient AML/CFT system. According to the 2007 IMF report, many of the reports filed were of poor quality (p. 168). In particular, the cash transaction reporting threshold triggering the identification obligation seemed problematic. An analysis of reporting patterns revealed that reporting entities tended to submit reports for transactions that exceeded this threshold without due diligence. On the other hand, transactions below the threshold were seemingly not properly scrutinised by financial institutions, as the low number of STRs referring to these transactions indicates (International Monetary Fund, 2007, p. 168). Therefore, in its 2007 review the IMF concluded that the over-reporting in Thailand could *inter alia* result from a misinterpretation of the reporting obligations (International Monetary Fund, 2007, p. 168).

Exposure of Reporting Entities

Another important element that could be noted in the countries with low reporting levels is the exposure of reporting entities filing the STR and the status of STR itself. Countries where reporting entities were excessively exposed to the attention of criminal authorities exhibited the tendency to have low reporting levels. In Liechtenstein the FIU was obliged to submit all suspicious activity reports to the prosecutor's office. This could negatively affect the willingness of financial institutions to file such reports (Moneyval, 2014, p. 12). Similar issues with the exposure of institutions obliged to reporting were also found in Austria. Under national law, the STRs constituted an instrument similar to a criminal complaint, meaning that the suspect or defendant had a right to access all the files including the relevant STRs after the end of proceedings. That made them a self-limiting

factor for reporting entities (FATF, 2014a, p. 12). A similar weakness was revealed in the German regime, where the reporting entity had to have a relatively high degree of certainty regarding the money laundering offense before filing a STR. Moreover, the reports had to be filed with the state criminal investigation authorities. As a result, the reports were perceived as criminal complaints rather than indicators of suspicious transactions (FATF, 2010, p. 12).

Secrecy provisions

Although bank and professional secrecy provisions in many jurisdictions have been removed or relaxed due to international pressure from the G20, the FATF and the OECD, they still seem to negatively impact the STR levels in several countries.

In 2014, it was underlined that Liechtenstein was still constrained by its national secrecy regulations. Austria also applied strict bank secrecy rules to information retained by banks. Moreover, there were restrictions on the FIU's access to information. However, following the legislative changes undertaken in 2010 and 2012, the follow-up report of 2014 up-graded Austria's legislation and considered it to be largely compliant with Recommendation 9 *"Financial Institution Secrecy Laws"* (FATF, 2014a, p. 5).

The 2010 MER criticized the German framework for having a carve-out provision for legal privilege and professional secrecy. Lawyers were exempted from their reporting obligations in instances where information on suspicious transactions was obtained while providing legal advice or while defending clients at court.[77] The negative implications of this exemption on reporting levels were enhanced by a broad interpretation of the legal advice definition. For instance, notaries claimed that almost all of their activities were considered to involve an element of legal advice, since attesting to the correctness of documents includes determining whether the terms of contracts and other documents are legally correct. The secrecy policy in Germany was further strengthened by the possibility to transmit STRs to competent federal chambers. Nevertheless, according to German chambers' representatives, this latter aspect did not play a significant role as STRs were forwarded promptly to the relevant authorities (FATF, 2010, p. 225).

This section has shown that differences in legal AML/CFT frameworks and the role of FIUs strongly affect STR levels. More specifically, the country studies have illustrated that (i) the scope of predicate offense under domestic laws, (ii) non-compliance with reporting obligations, (iii) the visits of evaluators and regulatory pressure, (iv) the efficiency of AML/CFT frameworks, (v) the exposure of reporting entities, and (vi) secrecy provisions in domestic law have an impact on STR levels. Acknowledging these very different legal frameworks, the next section broadens

[77] There was an exception to this regulation. The carve–out was not effective if it was certain that a client deliberately used legal advice for the purpose of ML or TF.

the perspective and analyses the effects of various economic and political factors that potentially drive STR levels.

Economic analysis of the drivers of STR levels

Introduction

This section aims at identifying factors that affect countries' STR in order to facilitate the determination of adequate reporting levels. Little is known about the determinants of reporting levels and empirical evidence on the effectiveness of STR schemes is rare. However, as there is broad consensus that money laundering and terrorist financing pose pressing political challenges, it is essential to establish reporting systems that facilitate effective prosecution. As the number of transactions reported varies significantly between countries and reporting obligations pose a substantial burden for financial and political institutions, a better understanding of the factors that drive reporting levels is much needed.

This section, therefore, provides an exploratory analysis of the effects of a broad range of political and economic factors on countries' reporting levels. Generally, a theoretical framework that allows evaluating the adequateness of reporting standards is missing and it seems questionable whether an effective reporting system is characterized by a high number of STRs. In a first step we thus develop a set of hypotheses and describe the data (Section 4.2). Subsequently, we regress national reporting levels on selected key factors in order to identify their respective effects on STR levels (Section 4.3). The findings are then discussed and summarized in Section 4.4.

Descriptive Data Analysis and Hypotheses Development

Since its founding in 1989, the FATF expanded to its current 36 members. It moreover cooperates closely with several associate members and observer organizations.[78] However, despite the FATF's general recommendation to publish suspicious activity reports annually, most financial intelligence units of associated member states do not provide comprehensive access to information on STRs. Our sample is thus restricted to countries that publish annual reports or grant access to data on reporting levels.[79]

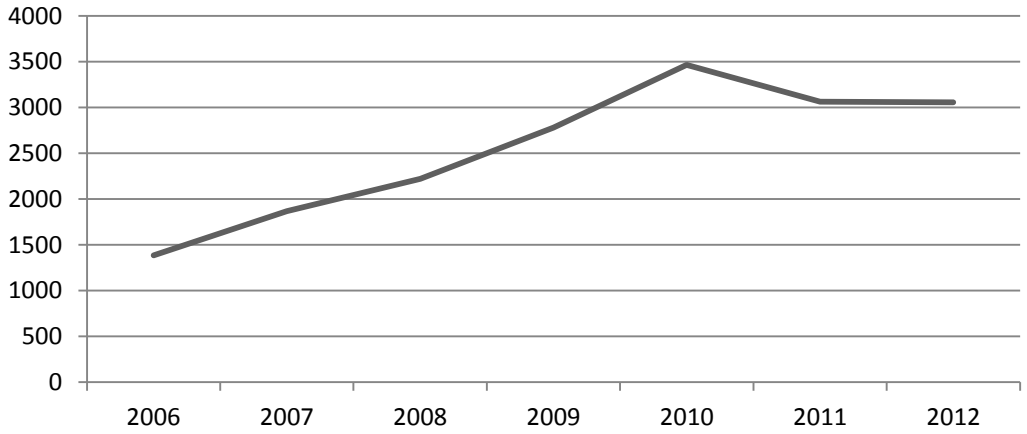
In total, we consider 54 countries[80] during the period from 2006 to 2012. Generally, countries' reporting levels vary dramatically. They range from 63 STRs in Malta (2009) to over 1.5 million SARs in the United States (2012).[81]

[78] http://www.fatf-gafi.org/about/membersandobservers/.
[79] These figures can for instance be retrieved from http://www.fiba.net/pages/FinancialIntelligenceUnits/.
[80] Countries that provided information on STRs are listed in Appendix A.
[81] Suspicious Activity Reports (SARs) are generally defined more broadly than STRs, as they include various activities that violate criminal law.

Figure 2 shows the total number of STRs (in thousands) over the period analysed. The number increased steadily from 1.38 million in 2006 to above 3.4 million in 2010. After this spike in 2010, we observe a small decrease in the following two years, with the total number of STRs remaining above three million. For 2012, our data comprises a total of 3.06 million reported STRs (SARs). With 1.58 million STRs, the lion's share of these STRs/SARs was reported by the United States (51.9% of total reports). Other important reporting countries were Japan, South Korea, and the United Kingdom, each with more than 250,000 STRs in 2012.

Clearly, the high overall number of STRs/SARs is driven by a small number of countries. In 2012, the first eight countries together reported more than 90% of all STRs. Therefore, it is imperative to investigate country-specific characteristics of the enforcement regime (see Section 3). In order to facilitate the comparison between different countries that adopted AML/CTF schemes, we additionally evaluate the effect of several macro-economic factors on reporting levels. Appendix B provides an overview of the variables considered in our analysis.

Figure 2: Total number of STRs (in thousands) from 2006 to 2012 in the sample of 54 countries



To start with, the economic size of a country seems an obvious determinant of varying STR levels among countries. Particularly the larger size of the financial sector may trigger more STRs. Indeed, a simple scatterplot illustrates that the number of STRs is positively correlated with the economic size of a country (measured in GDP) (Figure 3). Given this positive correlation, we "normalize" the number of STRs by the GDP of the respective country. We can thus analyse how specific factors affect the ratio of reports to GDP.

Figure 4 depicts the (natural logarithm of the) average ratio of STR/GDP between 2006 and 2012. Apart from a small decline in 2007, the average ratio of STR to GDP has increased steadily since 2006, exhibiting a relatively steep rise between 2007 and 2010 and a slower increase in the years after.

Figure 3: The number of STRs(ln) plotted against the GDP(ln) in the 54 sample countries between 2006 and 2012

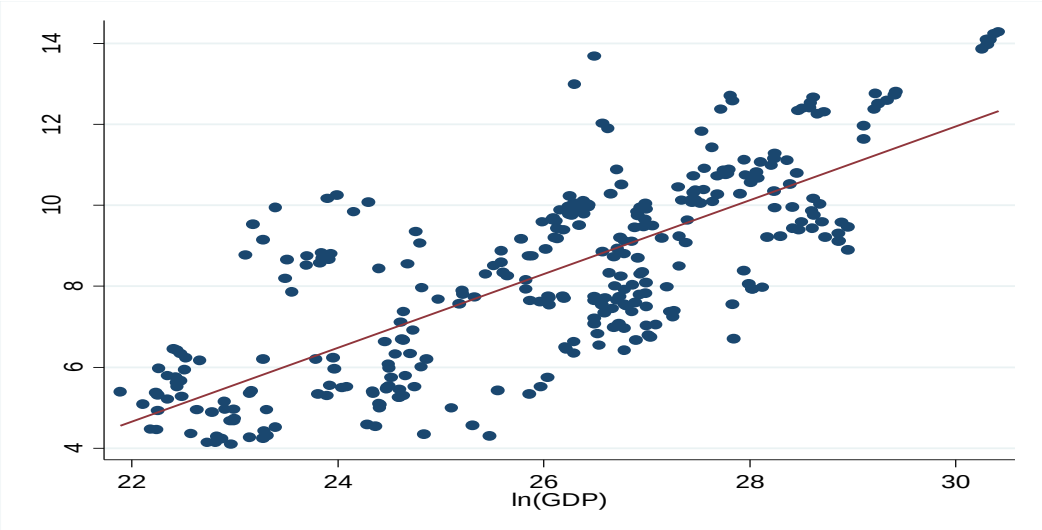
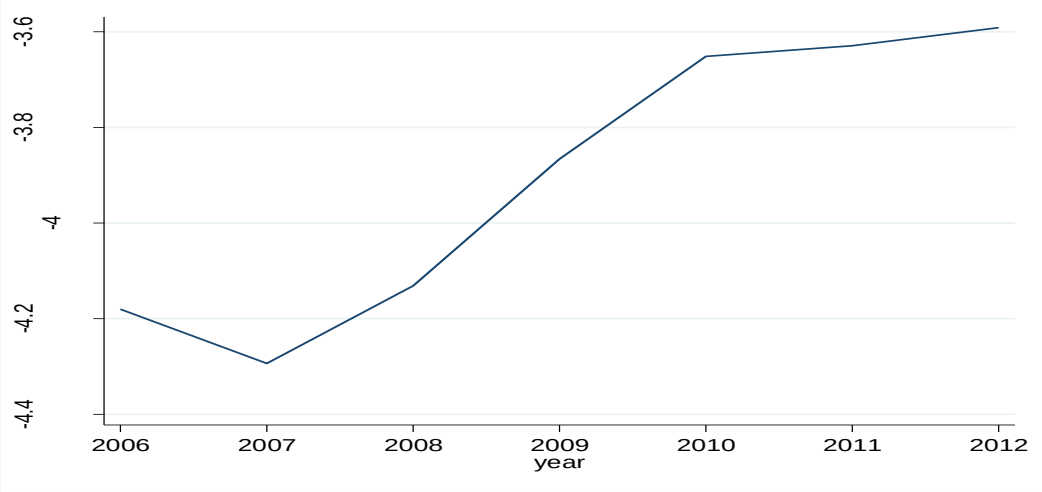


Figure 4: The average ratio of STR to GDP (ln) of the 54 sample countries between 2006 and 201



As little is known about the effects of those factors on countries reporting levels and it is often not clear how they affect STR, we develop a broad set of hypotheses that cover various economic and institutional factors.

First, we expect the level of economic development and general institutional factors to influence reporting levels. A state's stability, its capacity to fight corruption, and the size of the shadow economy likely determine suspicious activity. For instance, Puffer, McCarthy & Jaeger (2016) argue that institution building in emerging economies is key in fighting corruption, and Vaithilingam & Nair (2007) find less money laundering activity in countries with strong institutional capacity, effective legal and regulatory frameworks, and efficient governance structures. On the other hand, while a large degree of corruption, for instance, potentially triggers a lot of suspicious activity that translates into a high number of STRs, a corrupt system

might also undermine reporting standards and thus reduce the number of reports filed. Thus, we hypothesize that:

H1: General institutional factors affect STR levels (direction unclear).

Second, we assume that the depth and internationality of a country's financial sector impacts STR levels. We would expect that the larger this share of international activity in total banking activity is, the higher the risk for suspicious transactions taking place and being reported. The financial sector variables also capture that the place of financing of the criminal activity may differ from the place where the underlying criminal activity actually takes place.

H2: Depth and internationality of the financial sector increase the number of STRs.

Third, we conjecture that financial sector regulation influences STR levels. Restrictions on bank activity and the strength of official supervisory power, for instance, could have positive or negative effects on reporting levels. Therefore, we assume that

H3: Financial sector regulation affects STR levels (direction unclear).

Fourth, we expect that FATF membership increases STR levels. While little is known about the costs and benefits of AML/CFT regimes (Halliday, Levi & Reuter, 2014), FATF membership indicates a political commitment that might ultimately result in high numbers of STRs. In line with our findings from Section 3.4., we moreover expect on-site visits and MERs to increase STR levels, as they facilitate and supervise the implementation of FATF standards. It follows that

H4: FATF membership and mutual evaluations increase STR levels.

Fifth, we hypothesize that criminal activities drive reporting levels. Given that well-targeted reporting systems successfully identify criminal activities, countries with high crime-rates should file more STRs. Thus, particularly high costs for businesses that are caused by terrorism or organized crime are likely to increase the number of STRs. We therefore assume that

H5: Criminal activities increase the amount of STRs.

We finally conjecture, that tax havens oppose disclosure of financial information and thus report relatively few suspicious transactions. This is in line with the notion that tax havens protect investor privacy (Schjelderup, 2016). Thus, we hypothesize that

H6: Tax havens exhibit low STR levels.

Regression analysis

Data and methodology

We set up a panel data set comprising 54 countries and perform regression analysis to investigate the hypotheses above. The choice of countries depends mainly on whether or not the countries publish their STR numbers. The sample thus contains very different economies from all continents, including 28 OECD countries, a number of middle-income countries such as India, the Philippines or the Senegal, large countries such as the United States and very small countries such as Malta, and also nine jurisdictions which are classified as tax havens.[82] The full list of countries is included in Appendix A.

The sample period spans the years 2006 to 2012. The average number of reports per country over the sample period 2006-2012 amounts to 55,568 STRs per year. The average number of STRs per country exhibits a large variance ranging from a maximum of over 1.3 million in the United States and a minimum of 76 in Senegal. On average, there are 94 STRs per USD one billion GDP over all countries and all years in our sample. The highest ratios are to be found in Thailand (1,623.3), Latvia (864.1) and Georgia (759.3), while Qatar has the lowest ratio with 1.3, followed by Switzerland (1.9) (see Table A). Due to the limited availability of STR data we do not have observations for all countries for all years of the sample period. The panel is thus unbalanced.

To control for time-invariant country-specific factors (such as institutional and legal frameworks), fixed-effects estimations are used. These estimates, however, only give a picture on within-country variation over time. Hence, we additionally show random-effects estimates capturing some cross-country variation as well.

As clearly visible in Figure 1, the number of STRs shows a steady upward trend over most of the sample period. Also, our sample period includes the years of the financial and economic crisis starting in 2008 which might affect STR levels. All regression models thus include year-fixed effects to control for a common time-trend. Standard errors are clustered at the country level. The dependent variable is the (natural logarithm of the) ratio of the number of STRs over GDP by a given country in a given year. The choice of explanatory variables is based on the hypotheses above. Summary statistics and correlation tables are presented in Tables C and D in the Appendix.

[82] Cyprus, Hong Kong, Ireland, Lebanon, Luxembourg, Malta, Mauritius, Singapore, and Switzerland (Gravelle, 2015).

RESULTS

The baseline regression results with fixed effects are depicted in Table 1 below. Following hypothesis 1, we include a country's GDP per capita (gdppc) in order to measure its level of economic development. This variable is highly correlated with a number of institutional variables such as corruption levels, the strength of the rule of law, the size of the shadow economy and state stability (hypothesis 1).[83] We thus expect this variable to capture such institutional factors as well. However, against our expectations it is statistically not significant. This may be due to the ambiguous impact of institutional factors (captured by the gdppc variable) on reporting levels.

We further account for the development of the financial sector, which we assume to have a positive impact on STR levels (hypothesis 2). We use a proxy to measure the size and depth of the financial sector: domestic credit as a percentage of GDP (credit). This variable has the expected positive sign, but proves, however, not to be statistically significant. Potential variables to measure the internationality of the financial sector would be the in- and outflows of foreign portfolio investments. These variables are, however, highly correlated with the GDP per capita and are thus not included in the regression.[84] In line with hypothesis 3, we also include a proxy that captures the stringency of rules for the financial sector: an index measuring overall restrictions on banking activities (restrictiveness, columns (3) and (4)) and, alternatively, an index measuring the power of the supervisory authorities (supervision, columns (5) and (6)). Both are positive, but not statistically significant in our regressions.

Based on hypothesis 4, we wish to capture the effects of the institutional framework of the FATF. FATF membership and in particular on-site visits of a FATF committee and the subsequent publication of a MER may increase STR levels. The regressions thus also include dummy variables that take the value of one in the year an MER is published (mer) or in the following respective years (mer(t+1), mer(t+2), mer(t+3)). The dummies are individually and also jointly statistically significant and positive. It thus seems that an on-site visit and the publication of the MER do indeed increase the numbers of STRs significantly. The effect seems quite sizeable.

In the year following the publication of the report, the model predicts around a quarter more STRs, in the subsequent year a third more STRs than the country would have without an on-site visit and publication of a MER. In the third year the effect shrinks, but still amounts to 20%. Countries hence seem to react strongly to the MERs and increase their efforts to report more suspicious transactions (also see 3.4). However, the increase seems not to be persistent, but rather to decrease after the second year.

[83] Correlations are depicted in Appendix E.
[84] Correlations are depicted in Appendix E.

Table 1: Baseline Regressions: Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
gdppc (ln)	-0.776 (0.200)	-0.393 (0.549)	-0.465 (0.419)	-0.002 (0.998)	-0.083 (0.883)	0.550 (0.381)
credit	0.002 (0.448)	0.002 (0.509)	0.001 (0.829)	0.0001 (0.964)	0.0003 (0.924)	-0.0004 (0.872)
terrorism	-0.468** (0.020)		-0.588*** (0.003)		-0.515** (0.013)	
organized crime		-0.319* (0.051)		-0.341** (0.036)		-0.352** (0.039)
restrictiveness			-0.005 (0.917)	0.006 (0.918)		
supervision					0.029 (0.487)	0.048 (0.282)
mer			0.143** (0.040)	0.128* (0.085)	0.121* (0.058)	0.093 (0.140)
mer (t+1)			0.270*** (0.009)	0.256** (0.011)	0.247** (0.013)	0.227** (0.010)
mer (t+2)			0.343*** (0.002)	0.355*** (0.001)	0.325*** (0.005)	0.346*** (0.002)
mer (t+3)			0.187* (0.056)	0.211** (0.047)	0.189** (0.047)	0.223** (0.037)
observations	270	269	260	259	242	241
countries	54	54	51	51	47	47
R^2	0.28	0.27	0.37	0.34	0.34	0.34
F-test	3.02 (0.01)	3.70 (0.00)	4.20 (0.00)	4.77 (0.00)	4.79 (0.00)	6.27 (0.00)
Joint significance mer (t+1) mer (t+2) mer (t+3)			3.91 (0.01)	4.33 (0.01)	3.33 (0.03)	3.80 (0.02)

Notes: The dependent variable in all regressions is the natural logarithm of the ratio of STRs to GDP. All regressions include time FE and country FE, and a constant. Standard errors are clustered at the country level; p-values in parentheses; sample period: 2006-2012; *, **, and *** indicate significance at the 10%, 5%, and 1% level; the number of years and observations varies due to limited data availability.

Moreover, we also try to account for the criminal activity taking place in a country, in particular terrorism (columns (1), (3) and (5)) and organized crime (columns (2), (4) and (6)), as we would expect that the numbers of STR vary with the extent of these activities in a country (hypothesis 5). We proxy these two variables using a yearly survey carried out by the World Economic Forum, which is available since

2006.[85] The questions asked in the survey are "In your country, to what extent does the threat of terrorism [organized crime (mafia-oriented racketeering, extortion)] impose costs on businesses? The answer categories vary from "to a great extent [1] to "not at all" [7]. Higher numbers hence reflect a lower perceived threat of terrorism and organized crime respectively.Table 2 gives an overview on the top seven countries in our sample for which the survey indicates the highest/lowest costs for business due to terrorism and organized crime on average over time.

Table 2: Top seven countries with highest/lowest costs reported due to terrorism and organized crime

Indicator	Countries
highest costs due to terrorism	Lebanon, United States, Nigeria, the Philippines, Turkey, India, Mexico
lowest costs due to terrorism	Finland, Slovakia, Iceland, Austria, Hungary, Lithuania, Czech Republic
highest costs due to organized crime	South Africa, Argentina, the Philippines, Turkey, Poland, United States, Croatia
lowest costs due to organized crime	Iceland, Denmark, Finland, Singapore, Luxembourg, Norway, Austria

In the regressions, both variables show the expected sign and are statistically significant, indicating that in fact STR levels are associated with the extent of criminal activity taking place in a country. Even if the magnitude of the effects themselves are difficult to interpret, as terrorism and organized crime are measured as indices, and as it is not entirely clear what an increase in the index by one percentage point means in real terms, one can say that terrorism seems to be reflected in STR levels more strongly than organized crime. [86]

In addition to the fixed-effects estimations presented above, we also performed some random effects estimations (Table 3). Dismissing country-fixed effects allows capturing some cross-country variation and estimating the impact of factors that do not vary over time. Moreover, these regressions serve as robustness checks. The regressions in columns (1) to (4) predict similar outcomes when compared to the fixed-effects estimations above. Columns (5) and (6) show regression results including a tax haven dummy for the nine tax haven countries in our sample. In line with hypothesis 6, the results do indicate lower STR levels in these jurisdictions. The effect is however not statistically significant.

In order to further assess the robustness of our results we re-run all regressions without the US, which is by far the country with the most STRs/SARs reported. As

[85] http://www.weforum.org/reports?filter[type]=Competitiveness.

[86] For illustrative purposes, a decrease by one percentage point in terrorism levels (as depicted by the data) would for instance be equivalent to a decrease from the level in Nigeria (index=3.7) to the Indian level (index=4.7).

the results remain largely unchanged, we are confident that the identified variables actually affect STR levels (see Tables F and G in the appendix).

Discussion of the econometric results

Despite the fact that some authors question the reasonableness of AML/CFT regimes (e.g. Takats, 2011), our analysis suggests that they are effective in identifying terrorism and organized crime, as countries with higher crime rates file more STRs. Moreover, we find that FATF membership and mutual evaluations drive reporting levels. Assuming that peer pressure likely promotes compliance with FATF standards, this finding seems plausible: Countries that are labelled as non-compliant face reputational risk that might, for instance, discourage foreign investors. Somewhat surprisingly however, we do not observe significantly deviating reporting patterns for tax havens. While strict non-disclosure regulations and limited exchange of information seem to be reflected in low STR levels, this effect is not statistically significant.

The regression results indicate conditional correlations, meaning they capture the effect of a specific variable while controlling for the effects of the other covariates in the regression. However, interpreting the outcomes in a causal way seems problematic, as it is conceivable that for some variables the causality runs in the opposite direction or in both directions. For instance, a country that already records a high level of terrorism might attract additional illicit money flows, which finance more terrorist activity (which are reflected in additional STRs).

Also, measurement errors may lead to endogeneity concerns. In particular, the measurement of STRs seems problematic and the question arises as to whether STRs are actually comparable across countries. In other words, does an STR in the USA mean the same as an STR in Qatar? The legal analysis shows that due to the soft law nature of the FATF recommendations and their rather flexible implementation in the different countries the qualitative identity of STRs across countries is not given (see Sections 2 and 3).

These endogeneity concerns may potentially lead to biased estimators, meaning that our results should be interpreted with some caution. It is not clear which factors are indicators and which factors are drivers of STRs. This problem is common in the literature on money laundering and the shadow economy, because - as in our case - measurement is difficult and there is no theory that guides the interpretation (Schneider, Buehn & Montenegro, 2010). Nevertheless, we hope that the links we find between the different factors contribute to the literature and may serve as a primer on which further research can build.

Table 3: Baseline Regressions: Random Effects

	(1)	(2)	(3)	(4)	(5)	(6)
gdppc (ln)	-0.257	-0.145	-0.087	0.057	-0.086	0.057

	(0.396)	(0.653)	(0.764)	(0.856)	(0.768)	(0.854)
credit	0.003	0.002	0.001	0.001	0.001	0.001
	(0.229)	(0.233)	(0.573)	(0.552)	(0.583)	(0.569)
terrorism	-0.339*		-0.481***		-0.482***	
	(0.066)		(0.006)		(0.006)	
organized		-0.244*		-0.293**		-0.294**
crime		(0.099)		(0.048)		(0.047)
restrictiveness			-0.011	-0.004	-0.011	-0.004
			(0.805)	(0.936)	(0.804)	(0.939)
haven					-0.095	-0.018
					(0.867)	(0.971)
mer			0.130**	0.119*	0.130**	0.119*
			(0.045)	(0.097)	(0.046)	(0.098)
mer (t+1)			0.247**	0.237**	0.247**	0.238**
			(0.012)	(0.013)	(0.012)	(0.013)
mer (t+2)			0.334***	0.338***	0.334***	0.338***
			(0.001)	(0.000)	(0.001)	(0.000)
mer (t+3)			0.183**	0.194**	0.182**	0.194**
			(0.038)	(0.043)	(0.037)	(0.042)
observations	270	269	260	259	260	259
countries	54	54	51	51	51	51
R^2	0.27	0.27	0.36	0.34	0.36	0.34
Wald test	37.65	29.68	60.79	77.80	61.27	77.38
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Joint significance mer (t+1) mer (t+2) mer (t+3)			12.74	13.87	12.66	13.85
			(0.01)	(0.00)	(0.01)	(0.00)

Notes: The dependent variable in all regressions is the natural logarithm of the ratio of STRs to GDP. Random effects estimations, all regressions include time FE, and a constant. Standard errors are clustered at the country level. P-values in parentheses. Sample period 2006-2012. *, **, and *** indicate significance at the 10%, 5%, and 1% level. The number of years and observations varies due to the limited availability of some data.

CONCLUSIONS

In combining legal and economic methods this paper investigates the effects of various policies, institutional and legal factors on STR levels. As earlier research questions the effectiveness of AML/CFT frameworks in general (e.g. Takats, 2011), recently a lot of attention has been drawn to the discussion regarding which information high STR numbers actually conveys. And indeed, this question is difficult to answer. While high numbers of STRs ideally indicate high levels of

suspicious activity or efficient monitoring systems, they might also reveal poorly-targeted evaluation systems that provide incentives for over-reporting.

Among the most important features of legal systems that drive numbers of STRs is the scope of predicate offenses in national criminal law. Moreover, sanctions for non-compliance with national AML/CFT legislation affect reporting levels. Depending on their design, sanctions may deter or encourage the filing of STRs. At worst, strict sanctions may lead to the phenomenon of *"defensive reporting"* or, on the contrary, cause a lack of reporting due to disproportionately high penalties. Therefore, adequate penalties for non-compliance with AML/CFT standards seem to be essential in the establishment of efficient reporting structures (Takats, 2011).

At the same time it has to be noted that high levels of STRs are not necessarily proof of an effective AML/CFT regime. Over-reporting may frequently be a consequence of inefficient legal systems, overly broad definitions of predicate offenses, or poorly calibrated automated suspicious transaction identification systems. Additionally, overly severe penalties for non-compliance with reporting standards potentially trigger over-reporting. Thus, as demonstrated in the foregoing analysis, apparent over-reporting should be examined in detail in order to draw conclusions on the effectiveness of AML/CFT frameworks.

The economic analysis partly confirms our hypotheses. In particular, we find that countries in which businesses face high costs due to terrorism or organized crime file more STRs. The supervision and reporting system therefore appears to be successful in detecting suspicious financial transactions related to money laundering and terrorism. Moreover, we find that the mutual evaluation framework increases reporting levels sustainably: countries report significantly more suspicious activity in the years following an evaluation. This indicates that the on-going review process is crucial for the improvement of international compliance with the FATF Standard. In identifying deficiencies in countries' STR regimes, the assessors apparently initiate reforms that increase compliance with the FATF recommendations and drive reporting levels. Consequently, mutual evaluation schemes should incorporate risk considerations and target those countries that are particularly vulnerable to criminal activities related to money laundering and terrorist financing.

Regarding countries' economic and financial sector development, we do not observe any significant effects on STR levels. Also, we do not find any effects of financial sector regulation and, likewise, tax havens exhibit no divergent reporting patterns. However, as the results suggest that the global AML/CFT system successfully identifies activity linked to terrorism and organized crime, the capacity of regulatory and supervision regimes needs to be further strengthened in order to effectively and sustainably implement the FATF Recommendations (OECD, 2014).

The review of the AML/CFT frameworks in the selected countries is useful in providing evidence on a possible correlation between STR levels and certain characteristics of national legal and institutional frameworks. This analysis is not sufficient to provide conclusive evidence on the extent of particular legal and institutional factors' impact on reporting levels. Moreover, it does not evaluate the effectiveness of different STR regimes. However, it clearly demonstrates that certain characteristics of legal and institutional frameworks strongly affect the quantity of STRs.

Our analyses do not cover some factors that are clearly relevant in the field of terrorism financing and money laundering. These include, for instance, cyber-money such as bitcoins. As today only few AML/CFT regimes contain provisions on virtual currencies, FIUs should in the future be provided with efficient means to respond to the digitalization of economies.

REFERENCES

Allingham, M., & Sandmo, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics, 1(3-4),* 323-338.

Barth, J.R., Caprio, G., jr., & Levine, R. (2013). Bank Regulation and Supervision in 180 Countries from 1999 to 2011. *National Bureau of Economic Research Working Paper 18733*. doi: 10.3386/w18733

Chaikin, D. (2009). How effective are suspicious transaction reporting systems? *Journal of Money Laundering Control, 12(3),* 238-253. doi:10.1108/13685200910973628

Cindori, S. (2013). Money laundering: correlation between risk assessment and suspicious transactions. *Financial Theory and Practice, 37(2),* 181-206. doi:10.3326/fintp.37.2.3

EUROSTAT. (2013). *Money Laundering in Europe*. Luxembourg: European Union. doi: 10.2785/415

FATF. (2003). *The Forty Recommendations.* Paris: FATF.

FATF. (2005). *Summary of the Third Mutual Evaluation Report*. Paris: FATF.

FATF. (2008). *Mutual Evaluation Report, Anti-Money Laundering and Combating the Functioning of Terrorism. Mexico*. Paris: FATF.

FATF. (2009). *Rapport d'évaluation mutuelle. Rapport de suivi. Lutte contre le blanchiment de capitaux et le financement du terrorisme*. *Suisse*. Paris: FATF.

FATF. (2010). *Mutual Evaluation Report, Anti-Money Laundering and Combating the Functioning of Terrorism. Germany*. Paris: FATF.

FATF. (2012). *International Standards of Combating Money Laundering and the Financing of Terrorism and Proliferation, The FATF Recommendations.* Paris: FATF.

FATF. (2014a). *Follow-Up Report Mutual Evaluation Report. Austria*. Paris: FATF.

FATF. (2014b). *Follow-Up Report Mutual Evaluation Report. Mexico*. Paris: FATF.

Gravelle, J. (2015). *Tax havens: International Tax Avoidance and Tax Evasion*. Congressional Research Service. Retrieved from https://www.fas.org/sgp/crs/misc/R40623.pdf.

Halliday, T., Levi, M., & Reuter, P. (2014). *Global Surveillance of Dirty Money: Assessing Assessments of Regimes to control Money-Laundering and Combat the Financing of Terrorism*. Center on Law and Globalization. Retrieved from http://www.lexglobal.org/files/Report_Global%20Surveillance%20of%20Dirty%20Money%201.30.2014.pdf.

International Monetary Fund. (2007). *Thailand: Detailed Assessment Report on Anti-Money Laundering and Combating the Financing of Terrorism*. Washington, D.C.: International Monetary Fund.

Johnson, J., & Desmond Lim, Y. (2003). Money laundering: has the Financial Action Task Force made a difference? *Journal of Financial Crime, 10(1),* 7-22. doi:10.1108/13590790310808556

Johnston, R. B., & Carrington, I. (2006). Protecting the financial system from abuse: challenges to banks in implementing AML/CFT standards. *Journal of Money Laundering Control, 9(1),* 48 - 61. doi:10.1108/13685200610645210

KPMG. (2003). *Money Laundering: Review of the Reporting System*. KPMG. Retrieved from http://www.dematerialisedid.com/PDFs/kpmgreport.pdf.

MENAFATF. (2008). *Mutual Evaluation Report, Qatar*. MENAFATF.

MENAFATF. (2012). *Mutual Evaluation Report, Qatar*. MENAFATF.

Moneyval. (2001). *First Mutual Evaluation Report on Latvia, Summary*. Council of Europe.

Moneyval. (2006). *Anti-Money Laundering and Combating the Financing of Terrorism*. Council of Europe.

Moneyval. (2012). *Anti-Money Laundering and Combating the Financing of Terrorism*. Council of Europe.

Moneyval. (2014). *Anti-Money Laundering and Combating the Financing of Terrorism*. Council of Europe.

OECD. (2014). *Illicit Financial Flows from Developing Countries: Measuring OECD Responses*. Paris: OECD Publishing.

Puffer, S., McCarthy, D., & Jaeger, A. (2016). Institution building and institutional voids. *International Journal of Emerging Markets, 11*(*1),* 18-41.

Roule, T., & Kinsell, J. (2003). Legislative and bureaucratic impediments to suspicious transaction reporting regimes. *Journal of Money Laundering Control, 6(2),* 151-156.

Schjelderup, G. (2016). Secrecy Jurisdictions. *International Tax and Public Finance, 23(1),* 168-189.

Schneider, F., Buehn, A., & Montenegro, C. (2010). New Estimates for the Shadow Economies all over the World. *International Economic Journal, 24(4),* 443-461.

Takats, E. (2011). A Theory of "Crying Wolf": The Economics of Money Laundering Enforcement. *The Journal of Law, Economics & Organization, 27(1),* 32-78.

Terry, L. S. (2010). An Introduction to the Financial Action Task Force and its 2008 Lawyer Guidance. *Journal of the Professional Lawyer, 20*(*3),* 3-67.

Vaithilingam, S., & Nair, M. (2007). Factors affecting money laundering: lesson for developing countries. *Journal of Money Laundering Control, 10(3),* 352-366.

APPENDIX

A. List of Countries

Country	STR	STR/GDP	Country	STR	STR/GDP
Argentina	3,507	7.3	Luxembourg	4,205	76.3
Australia	36,590	32.0	Malta	92	11.0
Austria	1,859	4.6	Mauritius	158	17.0
Barbados	104	23.1	Mexico	49,085	43.2
Belgium	17,267	34.9	Netherlands	33,704	40.0
Bulgaria	746	15.2	New Zealand	4,249	34.2
Canada	57,393	35.7	Nigeria	3,188	8.0
Croatia	1,478	23.2	Norway	6,389	15.2
Cyprus	336	13.2	Philippines	9,506	45.1
Czech Republic	2,120	10.2	Poland	2,214	4.7
Denmark	1,828	5.6	Portugal	1,017	4.2
Estonia	5,552	254.7	Qatar	174	1.3
Finland	19,879	78.3	Republic of Korea	165,824	148.7
France	17,781	6.6	Romania	2,671	15.4
Georgia	9,995	759.3	Senegal	76	6.1
Germany	10,542	3.0	Serbia	2,449	61.8
Greece	1,702	5.0	Singapore	16,471	59.0
Hong Kong	17,739	78.2	Slovakia	2,344	25.9
Hungary	7,421	53.0	Slovenia	275	5.7
Iceland	377	24.1	South Africa	44,532	126.6
India	11,537	6.6	Spain	2,906	1.9
Ireland	12,488	56.3	Sweden	9,408	19.7
Italy	30,282	14.0	Switzerland	1,118	1.9
Japan	265,607	51.0	Thailand	495,018	1635.3
Latvia	24,304	864.1	Turkey	9,834	13.1
Lebanon	227	6.9	United Kingdom	249,031	95.6
Lithuania	213	5.2	United States	1,325,874	88.8

Note: STR and STR/GDP denote the average per country between 2006 and 2012 including all years where data are available. GDP is measured in billions USD.

B. Variables and Data Sources

Variable	Description and Source
gdp (ln)	Gross domestic product (GDP) measured in US dollars at current prices and current exchange rates. Source: World Bank Development Indicators
gdppc (ln)	GDP per capita measured in US dollars at current prices and current exchange rates. Source: World Bank Development Indicators
ofpi (ln) and ifpi (ln)	Outward and inward total portfolio investments, in US dollars at current prices and current exchange rates in millions. Source: FPI UNCTAD, UNCTADstat
state fragility index	Index combining scores on eight indicators measuring political and economic effectiveness and legitimacy. Index ranges from 0 "no fragility" to 25 "extreme fragility." Source: Systemic Peace Institute
claims(ln) and liabilities(ln)	Total Claims and Liabilities reported by banking offices located in the specified country regardless of the nationality of the controlling parent, in millions of US dollars; Source: Bank for International Settlements
credit	Domestic credit provided by financial sector (% of GDP). Source: World Bank: World Development Indicators
corruption	Freedom from Corruption Index; The index ranges from 0 to 100, with higher values indicating lower levels of corruption. Source: Heritage Foundation
shadow economy	Size of the shadow economy (in % of GDP) available for 1999-2007. Source: Schneider, Buehn & Montenegro (2010).
rule of law	Percentile rank terms from 0 to 100, with higher values corresponding to better outcomes.[87] Source: World Bank: Worldwide Governance Indicators
restrictiveness	Index ranging from 3 to 12, where higher values indicate more restrictive bank regulation.[88] Source: World Bank Survey (Barth, Caprio & Levine, 2013)
supervision	Index ranging from 0 to 14, where higher values indicate greater supervisory power.[89] Source: World Bank Survey (Barth, Caprio & Levine, 2013)
terrorism	Cost for Businesses caused by terrorism Survey question: In your country, to what extent does the threat of terrorism impose costs on businesses? [1 = to a great extent; 7 = not at all] Source: World Economic Forum: Global competitiveness index
organized crime	Cost for Businesses caused by organized crime Survey question: In your country, to what extent does organized crime (mafia-oriented racketeering, extortion) impose costs on businesses? [1 = to a great extent; 7 = not at all] Source: World Economic Forum
haven	Binary variable taking the value 1 if a country is classified as a tax haven and 0 otherwise. Source: Gravelle (2015).
mer	Binary variable taking the value 1 if an MER is published on a country in a specific year (0 otherwise); Source: http://www.fatf-gafi.org/

[87] Index captures perceptions of the extent to which agents have confidence in and abide by the rules of society and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence.

[88] Sum of three indices: The extent to which banks may engage in underwriting, brokering and dealing in securities, and all aspects of the mutual fund industry. + The extent to which banks may engage in insurance underwriting and selling. + The extent to which banks may engage in real estate investment, development and management.

[89] Whether the supervisory authorities have the authority to take specific actions to prevent and correct problems.

C. Summary statistics

	N	Mean	Std. Dev.	Min	Max
str/gdp (ln)	270	-3.97	1.59	-7.35	1.01
gdppc (ln)	270	9.91	1.08	6.85	11.64
credit	270	127.48	69.81	18.80	349.03
supervision	242	10.79	2.48	3.00	14.69
restrictiveness	260	6.73	1.70	2.25	12.00
terrorism	270	5.67	0.77	2.60	6.80
organized crime	269	5.61	0.91	2.69	6.85
mer	260	0.18	0.39	0.00	1.00
mer (t+1)	260	0.19	0.39	0.00	1.00
mer (t+2)	260	0.17	0.38	0.00	1.00
mer (t+3)	260	0.27	0.44	0.00	1.00
haven	260	0.18	0.39	0.00	1.00

D. Pairwise correlations

	str/gdp (ln)	gdppc (ln)	credit	super-vision	restric-tiveness	terrorism	organized crime	mer	mer (t+1)	mer (t+2)	mer (t+3)	haven
str/gdp (ln)	1.00											
gdppc (ln)	0.06	1.00										
credit	0.15	0.50	1.00									
super-vision	-0.02	-0.02	-0.08	1.00								
restric-tiveness	-0.08	-0.29	-0.07	0.38	1.00							
terrorism	-0.05	0.24	-0.11	-0.03	-0.13	1.00						
organized crime	-0.07	0.52	0.24	-0.06	-0.26	0.52	1.00					
mer	-0.02	0.05	0.02	-0.01	-0.10	0.02	-0.04	1.00				
mer (t+1)	0.02	0.07	0.04	0.02	-0.07	0.03	0.02	-0.17	1.00			
mer (t+2)	0.08	0.01	0.02	0.06	-0.05	0.01	0.02	0.01	-0.16	1.00		
mer (t+3)	0.12	0.08	0.09	0.11	-0.01	-0.01	0.04	0.01	-0.01	-0.11	1.00	
haven	0.04	0.24	0.22	0.03	-0.06	0.08	0.32	-0.01	-0.03	-0.03	-0.04	1.00

E. Correlations with GDP per capita

	gdppc (ln)	corruption	rule of law	state fragility index	shadow economy	foreign portfolio investment flows (ln)
gdppc (ln)	1.00					
corruption	0.86	1.00				
rule of law	0.81	0.96	1.00			
state fragility index	-0.75	-0.72	-0.72	1.00		
shadow economy	-0.77	-0.76	-0.72	0.34	1.00	
foreign portfolio investment flows (ln)	0.61	0.55	0.49	-0.28	-0.64	1.00

F. Regression Results without the United States (Fixed Effects)

	(1)	(2)	(3)	(4)
gdppc (ln)	-0.812	-0.446	-0.487	-0.047
	(0.188)	(0.500)	(0.408)	(0.942)
credit	0.002	0.002	0.001	0.0001
	(0.478)	(0.554)	(0.836)	(0.976)
terrorism	-0.464**		-0.595***	
	(0.023)		(0.003)	
organized crime		-0.347**		-0.363**
		(0.034)		(0.030)
restrictiveness			-0.005	0.010
			(0.914)	(0.854)
mer			0.139**	0.111
			(0.043)	(0.117)
mer (t+1)			0.271***	0.239**
			(0.010)	(0.014)
mer (t+2)			0.346***	0.334***
			(0.003)	(0.002)
mer (t+3)			0.185	0.192
			(0.105)	(0.118)
R^2	0.28	0.28	0.37	0.35
observations	263	262	253	252
countries	53	53	50	50
F-test	3.01	3.95	3.99	4.81
	(0.01)	(0.00)	(0.00)	(0.00)
Joint			3.46	3.58
significance			(0.02)	(0.02)
mer (t+1)				
mer (t+2)				
mer (t+3)				

Notes: The dependent variable in all regressions is the natural logarithm of the ratio of STRs to GDP. All regressions include time FE and country FE, and a constant. Standard errors are clustered at the country level; p-values in parentheses; sample period: 2006-2012; *, **, and *** indicate significance at the 10%, 5%, and 1% level; the number of years and observations varies due to limited data availability.

G. Regression Results without the United States (Random Effects)

	(1)	(2)	(3)	(4)
gdppc (ln)	-0.268 (0.380)	-0.155 (0.632)	-0.095 (0.746)	0.043 (0.892)
credit	0.002 (0.258)	0.002 (0.290)	0.001 (0.589)	0.001 (0.601)
terrorism	-0.321* (0.089)		-0.477*** (0.008)	
organized crime		-0.254* (0.091)		-0.300** (0.049)
restrictiveness			-0.012 (0.792)	-0.002 (0.964)
mer			0.128** (0.046)	0.105 (0.130)
mer (t+1)			0.248** (0.011)	0.224** (0.017)
mer (t+2)			0.336*** (0.001)	0.321*** (0.001)
mer (t+3)			0.181* (0.072)	0.178 (0.106)
R^2	0.27	0.27	0.36	0.34
observations	263	262	253	252
countries	53	53	50	50
Wald-test	29.55 (0.00)	39.23 (0.00)	58.55 (0.00)	77.22 (0.00)
Joint significance mer (t+1) mer (t+2) mer (t+3)			11.55 (0.01)	11.73 (0.01)

Notes: The dependent variable in all regressions is the natural logarithm of the ratio of STRs to GDP. Random effects estimations, all regressions include time FE, and a constant. Standard errors are clustered at the country level.
P-values in parentheses. Sample period 2006-2012. *, **, and *** indicate significance at the 10%, 5%, and 1% level. The number of years and observations varies due to the limited availability of some data.

Behavioural Economics and Tax Evasion: Calibrating an Agent-based Econophysics Model with Experimental Tax Compliance Data

Cécile Bazart[90]*, Aurélie Bonein*[91]*, Sascha Hokamp*[92]*, and Götz Seibold*[93]

Abstract

We observe in the literature a persistent lack of calibrating agent-based econophysics tax evasion models. However, calibrations are indispensable to the quantitative and predictive application of such computational simulation approaches. Therefore, we analyse individual data from two tax compliance experiments with social interaction: from information on tax enforcement measures in groups with income heterogeneity, where the audit probability is known and audit results are publicly and officially announced; and from information about the mean reported income of other group members in the previous period. In our agent-based econophysics simulation, we implement recent advances in behavioural economics, for instance to describe social interactions within a population of behaviourally heterogeneous taxpayers. For this purpose, we employ experimental data showing a bimodal distribution which allows us to apply Ising's description of magnetism, a model adopted from statistical physics that can be related to binary choice models. We restrict agents in our econophysics framework to show selfish, imitating, ethical or random motives in their decisions to declare income. We find that the subjects in the experimental laboratory pursue rather mixed behaviour, including random and imitating motives.

JEL Classifications: C63; C92; H26; O17

Keywords: Tax Evasion, Tax Compliance Experiments, Agent-based Model, Behavioural Economics, Econophysics, Calibration

Acknowledgments: We thank two anonymous reviewers for constructive comments and James Alm, Betty R. Jackson and Michael McKee for providing the raw data from their tax compliance experiment. However, all errors remain our own.

INTRODUCTION

Scholars of various disciplines contribute to behavioural economics, from social scientists to physicists. Their contributions question neoclassical assumptions such as, for instance, that subjects in the experimental laboratory do not always act as

[90] LAMETA, University of Montpellier 1, UFR Economie, Espace Richter, Avenue Raymond Dugrand, CS 79606, 34960 Montpellier Cedex 2, France, Phone: +33 434 432493, Email: cecile.bazart@lameta.univ-montp1.fr

[91] CREM, University of Rennes 1, Department of Economics, 7 Place Hoche, CS 86514, 35065 Rennes Cedex, France, Phone: +33 223 233540, Email: aurelie.bonein@univ-rennes1.fr

[92] Corresponding author, CEN, University of Hamburg, Research Unit Sustainability and Global Change, Grindelberg 7, 20144 Hamburg, Germany, Phone: +49 40 428382053, Email: sascha.hokamp@uni-hamburg.de.

[93] Brandenburg University of Technology Cottbus-Senftenberg, Faculty 1, Institute of Physics, PO Box 10 13 44, 03013 Cottbus, Germany, Phone: +49 355 693006, Email: goetz.seibold@b-tu.de.

rationally and selfishly as taxpayers in the expected utility models of Allingham and Sandmo (1972), Srinivasan (1973) and Yitzhaki (1974).[94] Of course, in behaviourally heterogeneous populations, the tax evasion decision is embedded in a highly complex system of social interactions. Such complex systems are especially appropriate to a non-standard computational approach called agent-based modelling. This approach often succeeds in providing predictions that match real-life observations, because various kinds of interaction between autonomous agents are a common feature of agent-based models (Macal & North, 2005; Rand & Trust, 2011).

We employ statistical mechanics in an economic context to model social interactions via Ising's description of magnetism (Ising, 1925), an approach belonging to *econophysics* that combines economics and physics.[95] In this paper, closely related to Train's (2009) and Sornette's (2014) binary discrete choice models, we apply an econophysics approach with dual aims: to analyse tax evasion behaviour, and to provide a first attempt to calibrate our agent-based econophysics model with experimental tax compliance data (Alm, Jackson, & McKee, 2009; Bazart & Bonein, 2014). Having described the theoretical model, we make use of these experimental data to test its reliability in terms of the adequacy of its theoretical and empirical findings and show its flexibility in terms of predictions, for example to identify parameter settings of interest to future experimental research. In line with previous tax compliance experiments, we confirm that social networks play an essential role in individual decisions on income declarations, and find that the majority of subjects in the experimental laboratory show a complex pattern of attitudes, mixing selfishness or ethics with imitation.

The remainder of this paper proceeds as follows. In the next section, we provide a literature overview focusing on the calibration of agent-based models of tax evasion. Next, we present our agent-based econophysics tax evasion model, including the types of behavioural agents implemented. We then briefly introduce the experimental designs of the studies used and their findings, and perform calibrations of our agent-based econophysics model with their experimental data, before making some concluding remarks.

Literature review

Agent-based tax evasion models can be categorized into economics and econophysics branches (Hokamp & Pickhardt, 2010).[96] Zaklan, Lima, and

[94] For a survey of tax compliance experiments in behavioural economics, see Alm (2010).
[95] Schulz (2003) and Stauffer (2013) provide literature reviews of econophysics. Schinckus (2013) distinguishes in his overview between statistical and agent-based econophysics.
[96] Our literature review supplements four surveys by: (i) Bloomquist (2006) on three early agent-based frameworks for tax compliance (Mittone & Patelli, 2000; Davis, Hecht, & Perkins, 2003; Bloomquist, 2004a, 2004b, 2008); (ii) Alm (2012) on recent advances in tax evasion from theory, experiments and field studies; (iii) Hokamp (2013) on agent-based tax evasion modelling; and (iv) Pickhardt and Prinz (2014) on the behavioural dynamics of tax declarations.

Westerhoff (2008) and Zaklan, Westerhoff, and Stauffer (2009) launched the econophysics branch, based on Ising's description of magnetism (Ising, 1925).[97] One finding of these early econophysics models is that enforcement always triggers tax compliance behaviour, regardless of the prevailing social network structure. In their seminal paper, Zaklan, Westerhoff, and Stauffer (2009) find rather low rates of tax evasion for high audit probabilities. Adding the majority-vote-model and an Apollonian network, Lima (2010, 2012a, 2012b) shows the robustness of Zaklan, Westerhoff, and Stauffer (2009), thereby providing an agent-based replication study in the field of tax evasion.

Seibold and Pickhardt (2013), Hokamp and Seibold (2014b) and Pickhardt and Seibold (2014) use an agent-based econophysics approach to tax evasion based on Zaklan, Westerhoff, and Stauffer (2009) and Hokamp and Pickhardt (2010). Pickhardt and Seibold (2014) successfully replicate both underlying settings, and thus link the econophysics and economics branches of agent-based tax evasion frameworks. Seibold and Pickhardt (2013) conclude that, *ceteris paribus*, increasing the number of tax-relevant periods subject to back auditing helps to reduce tax evasion. Hokamp and Seibold (2014b) find that higher levels of public goods provision increase tax compliance. Finally, Crokidakis (2014) employs an econophysics three-state kinetic opinion exchange model to show that, above a critical threshold for the coupling of agents, tax enforcement successfully combats tax evasion. However, these econophysics studies are calibrated with neither empirical nor experimental data.[98] Hence, we continue with a review of calibration attempts in the economics branch.

To the best of our knowledge, Bloomquist (2011a) was the first to provide a calibration of an agent-based tax evasion model. In particular, his calibration employs data from the National Research Program (NRP) of the Internal Revenue Service (IRS) for the 2001 tax year, as well as tax compliance experiments, and presents strong evidence that the attitudes to risk aversion of subjects in the experimental laboratory are similar to those of small businesses in agent-based computational simulations. Arsian and İcan (2013a, 2013b) build on Bloomquist (2011a) to conduct a tax evasion analysis for Turkey, calibrated with data from annual reports of the Turkish Revenue Administration. The authors find that von Neumann and Moore neighbourhoods are the essential social network structures to reduce tax evasion behaviour. Bloomquist (2011b) studies a synthetic county and concludes that mixed interactive auditing of heterogeneous agents is more effective than random audit strategies. Bloomquist and Koehler (2015) employ and calibrate Bloomquist (2011b), using artificial taxpayer data from Bloomquist (2012). Testing

[97] Note that punishment in agent-based econophysics tax evasion models refers to Davis et al. (2003), in particular the notion of penalization through pre-announced time periods in which a detected tax evader has to be fully tax compliant.

[98] Hokamp and Seibold (2014a) use aggregated experimental tax compliance data (Bazart & Pickhardt, 2011) to calculate that France seems to have a larger fraction than Germany of subjects rationally engaged in the shadow economy. Thus, they provide a calibrated agent-based econophysics model of the shadow economy.

four audit strategies, the authors show that, in terms of reducing misreported taxes, the most effective strategy is to ensure a minimum level of audits for each class of taxpayers.

Nordblom and Žamac (2012) utilize a survey of black-market service purchase in Sweden to confirm that the elderly evade substantially less tax than younger people. Miguel, Noguera, Llácer, and Tapia (2012) develop an agent-based tax evasion model for Spain to investigate behavioural mechanisms. Based on the latter model, Llácer, Miguel, Noguera, and Tapia (2013) find that considering only rational agents overestimates tax evasion, whereas social interaction allows the generation of more plausible tax compliance levels. Furthermore, Noguera, Llácer, Miguel, and Tapia (2014) calibrate Miguel et al.'s (2012) and Llácer et al.'s (2013) agent-based framework with empirical data from Spain. The authors conduct computational simulation experiments and find that social norms do not always optimize tax compliance. Garrido and Mittone (2013) calibrate their agent-based model on tax compliance data from Chile and Italy. Given income inequality, the authors find that tax authorities may optimize tax collection by auditing taxpayers who behave more frequently according to the bomb crater effect (Krauskopf & Prinz, 2011).

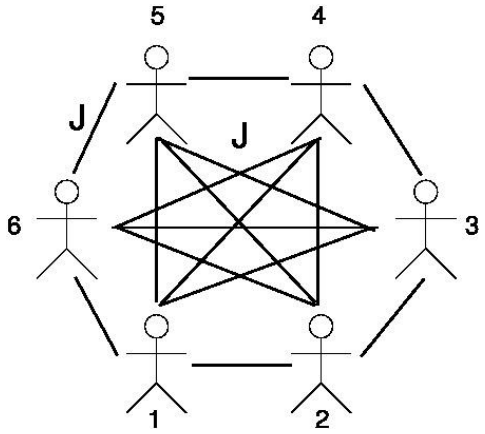
To summarize, these contributions support the relevance of modelling tax evasion decisions and social interactions in complex environments. However, we underline a persistent lack of agent-based econophysics tax evasion models regarding the purpose of calibrating computational simulations with empirical or experimental data. Such calibration might reveal both the theoretical validity and the predictive power of this tool. In the above review, we have identified six calibrated agent-based tax evasion frameworks: (i) Bloomquist (2011a) with IRS NRP and experimental data; (ii) Bloomquist (2011b) and Bloomquist and Koehler (2015) with artificial and IRS NRP data; (iii) Nordblom and Žamac (2012) with survey data from the Swedish Tax Agency; (iv) Miguel et al. (2012), Llácer et al. (2013) and Noguera et al.(2014) with Spanish empirical data; (v) Arsian and İcan (2013a, 2013b) with data from the Turkish Revenue Administration; and (vi) Garrido and Mittone (2013) with experimental data from Chile and Italy. In contrast to these calibrations based on the economics branch of tax evasion simulation models, our aim is to calibrate an agent-based econophysics model, which we present in the next section.

The agent-based econophysics approach

Within our theoretical framework, we adopt a simplified perspective on the description of tax evasion, namely that taxpayers are agents who choose between two alternatives: to declare either all or zero income to the tax authorities. The reduction of a continuous variable (i.e. the income declaration) to a binary variable may seem a drastic simplification; however, we demonstrate below that this kind of behaviour is frequently found in tax compliance experiments, and even emerges

in data from the IRS NRP for small business filers in tax year 2001 (Bloomquist, 2011a). Hence, our formal description within the Ising model, adopted from physics, corresponds with a model of discrete choice, the so-called logit binary choice model (Train, 2009; Sornette, 2014). The econophysics formulation has the advantage of providing a simpler theory structure, especially for the case of interacting agents, i.e. taxpayers. In any case, all physical quantities that appear in this model have a one-to-one correspondence in the economic language, as will be detailed below.

Figure 1. Sketch of a network considered within our econophysics approach



Note: With reference to Bazart and Bonein's (2014) tax compliance experiment, each group consists of N = 6 agents which are mutually connected by an exchange coupling, J.

The Ising-model Hamiltonian

$$(1)\ \mathrm{H} = -J \sum_{ij} S_i S_j - \sum_i B_i S_i$$

describes the coupling of Ising-variables (spins) Si = ±1 between group members (labelled with i = 1, . . . , N). For instance, Figure 1 shows the social network for N = 6 agents (Bazart & Bonein, 2014). The coupling strength, J is taken as a constant between group members, and we note that each pair (ij) is only counted once.

In the present context, we interpret $S_i = +1$ ($S_i = -1$) as a compliant (non-compliant) agent. Equation (1) also contains the coupling of the spins with a local magnetic field B_i, which may be associated with agents' moral attitudes.[99] In addition, our econophysics model contains a local temperature, T_i which measures the susceptibility of agents to external perturbations (either influence of neighbours or magnetic field). We then use the heat-bath algorithm to evaluate the statistical averages of the model (Krauth, 2006). The probability of a spin at lattice site *i* taking values $S_i = \pm 1$ is given by

[99] Note that our modelling of moral attitudes corresponds with parameter γ_i in Nordblom and Žamac's (2012) agent-based theory.

$$(2) \qquad p(S_i) = \frac{1}{1+\exp\{-[E(-S_i)-E(S_i)] \,/\, T_i\}}$$

and $E(-S_i) - E(S_i)$ is the energy change for a spin-flip at site *i*. On picking a random number $0 \leq r \leq 1$, the spin takes the value $S_i = +1$ when $r < p_i$ $(S_i = +1)$, and $S_i = -1$ otherwise. Obviously, one tax-relevant period then corresponds with a sweep through all members of all categories or groups.

Equation (2) has the same form as the decision probability in the logit discrete choice model, which allows for a mutual mapping of the corresponding quantities. In particular, by rewriting Equation (1) in the form

$$(3) \qquad H = -\sum_i \{J \sum_j S_j + B_i\} S_i = \sum_i E(S_i)$$

it turns out that the energy $E(S_i)$ of the 'Ising' system corresponds with the negative observable part of the utility function for the agent at site *i*. This agent will choose the alternative which maximizes her utility (i.e. lowers the energy). This utility function has two contributions, a term $\sim B_i$ which reflects the endogenous (moral) attitude of the agent towards evasion, and a second term $\sim J \, \Sigma_j \, S_j$ which captures the influence of the agent's social environment. The utility function is then maximized as the agent gets closer to the declaration behaviour of her neighbours in the network. Moreover, the temperature parameter T_i can, again by analogy with logit discrete choice, be interpreted as the standard deviation of the unobserved utility part corresponding with the spread in the non-measurable taste or attitude. Table 1 summarizes the parameters of the agent-based econophysics tax evasion model and compares their interpretation within the physical and economic contexts.

We then implement an enforcement scheme in our model, reflecting a case where the detection of an evading agent enforces tax compliance over the following *h* tax-relevant periods (or time steps). Zaklan, Lima, and Westerhoff (2008), Zaklan, Westerhoff, and Stauffer (2009), Lima (2010), Pickhardt and Seibold (2014) and Hokamp and Seibold (2014a, 2014b) invoke such a procedure, whereas Lima and Zaklan (2008) implement a randomized variant.

Table 1. Parameters of econophysics model and interpretation in physical and economic contexts

Variable	Physical Meaning	Economic Interpretation
S_i	Magnetic Moment at Position *i*	Decision Alternatives of Agent *i*: $S_i = +1$: compliant $S_i = -1$: non-compliant
B_i	Magnetic Field at Position *i*	Parameterized Moral Attitude of Agent *i* $B_i > 0$: Moral Behaviour $B_i < 0$: Amoral Behaviour
T_i	Local Temperature at Position *i*	Variance in Attitude of Agent *i*
J	Exchange of Energy between Magnetic Moments	Social Interaction Parameter between Agents
$E(S_i)$	Effective Energy for Spin at Position *i*	Negative Observable Part of the Utility Function for Agent *i*
H	Total Energy of the System	Negative Observable Part of the Aggregated Utility Function

Note: Parameters of the econophysics model described in Equations (1) and (3), and their interpretation within the logit discrete choice model applied to a binary model of tax evasion. The interaction parameter J is set to $J \equiv 1$, and therefore defines the scale for all other parameters.

Furthermore, Seibold and Pickhardt (2013) study generalizations of the auditing scheme to include time lapse effects. We set our audit probability to $p_a = 2/5$ and $p_a = 1/3$, corresponding with the values used in the tax compliance experiments used to calibrate our model (Alm, Jackson, & McKee, 2009; Bazart & Bonein, 2014).

Based on Seibold and Pickhardt (2013), Pickhardt and Seibold (2014) and Hokamp and Seibold (2014a,b), and following Hokamp and Pickhardt (2010), we assume that taxpayers can be classified into four categories of agent: (1) selfish a-type agents, which take advantage of non-compliance and are thus modelled via the parameter ratios $|B_i|/T_i \gg 1$ and $|B_i|/J \gg 1$ with $B_i < 0$; (2) imitating b-type agents, which conform to the norm of their social network, which in the model is realized through $B_i = 0$ and $J/T_i \gg 1$; (3) ethical c-type agents, which have large moral doubts and thus are practically always compliant, with behaviour parameterised by $|B_i|/T_i \gg 1$ and $|B_i|/J \gg 1$ and $B_i > 0$; and (4) random d-type agents, which act by chance, within a certain range, due to confusions caused by tax law complexity, with behaviour modelled by $B_i = 0$ and $J/T_i \ll 1$.

In the next section, we present the two tax compliance experiments (Alm, Jackson, & McKee, 2009; Bazart & Bonein, 2014) used with a view to exploring the composition of these four behavioural categories of agents.

Experimental design

In this section, we present the tax compliance experiments used to calibrate our econophysics model, the main parameters of which are reported in Table 2.

Table 2. Experimental settings used for our calibrations

Source	Alm, Jackson, & McKee (2009)	Bazart & Bonein (2014)
Treatment	Official Information (T2A)	Horizontal Inequity (H-I)
Income	Heterogenous: 100, 90, 80, 70, 60	Homogenous: 100
Tax Rate	35%	30%
Audit Probability	2/5	1/3
Fine	150%	350%
Auditing Information	Provided	Not Provided
Social Information	Not Provided	Provided
Groups	5	8
Group Size	8	6
Rounds	15	20
Total Number of Observations	600	960

Note: Auditing information means public and official announcements of audit results. Social information reflects the individual knowledge of the mean reported income of other group members.

Since social interaction is the crucial ingredient of our agent-based econophysics tax evasion simulation, it is necessary to focus on tax compliance laboratory experiments allowing social interactions. To this end, we use experimental data from Alm, Jackson, and McKee's (2009) study, which allows social information by

providing information on audits, and data from Bazart and Bonein (2014), in which social interactions are introduced through the provision of information on the average declaration of other group members.

Alm, Jackson, and McKee (2009) analysed the effect on tax compliance behaviour of dissemination of information on audit frequency and results. For this purpose, they implemented a pure declaration game in which subjects first performed a real-effort task in order to earn their income.[100] Next, they had to report their income for taxation at a 30 per cent rate. Because earned income was private information, any of them could under-report and decrease their tax burden. A random audit procedure was thus implemented to detect evasion that might result in the reimbursement of due tax plus payment of a penalty at 150 per cent. Subjects were placed in groups of six or eight individuals, but they did not know with whom they were playing during the 30 periods of the declaration game. To avoid cross-effects, redistribution was excluded, and to avoid end-of-game effects, participants were not informed of the exact number of periods in the declaration game. Alm, Jackson, and McKee (2009) implemented six treatments in a between-subjects design that differed depending whether and what type of information on audits was provided to the subjects. In the first three treatments, the audit probability was known to the subjects (Case A). These treatments differed in the announcement of audit results (no public announcement in T1A and T3A; public announcement in T2A) and unofficial communication (no communication in T1A and T2A; communication in T3A). The remaining three treatments were symmetric (Case B) but did not allow for an announcement of the audit probability. Unofficial communication was organized by allowing participants to send one message per round to all members of their group, mentioning whether they had been audited or caught cheating. Subjects' earnings were evaluated using all periods of the tax declaration game. The results support the positive effect of information on deterrent tools, more specifically when subjects have prior knowledge of the audit probability. For our purpose of calibrating an agent-based econophysics tax evasion model, we restricted ourselves to using the data in treatment T2A because this setting provided the maximum official information: both the audit probability and the audit results were announced. In addition, we used only the declaration choices made for each period by subjects facing an audit probability equal to $p_a = 2/5$. This reduced the sample to 40 subjects (five groups of eight subjects) for 15 periods, resulting in 600 declaration decisions.[101]

The second set of experimental data was drawn from Bazart and Bonein's (2014) study that introduced social interaction between subjects through the provision of information on the average declaration of other group members. The benchmark treatment in Bazart and Bonein (2014) was a pure declaration game, excluding redistribution through the provision of public goods financed by tax payments. In

[100] The earning procedure generated heterogeneity in income.
[101] The audit probability changed once at period 16.

this way, redistribution outcomes could not influence conditional reactions to others' declaration decisions. Groups of six subjects were formed and the group composition remained constant throughout the tax game. To avoid complex comparisons, all members of a group had the same income and faced the same fiscal policy parameters (i.e. tax, audit and penalty rates). At the beginning of each tax-relevant period, subjects were presented with a screen informing them of their individual income and the tax policy parameters, which were set such that they delivered the theoretical predictions of full compliance. At the beginning of each period, subjects received a constant income of 100 points each and faced: (i) a penalty rate of 350 per cent (including reimbursement of due taxes plus the fine); (ii) an audit probability of pa = 1/3 (audits were random and perfect); and (iii) a benchmark tax rate of 30 per cent. At the time when they made their decisions, the subjects had to determine the amount of income they would self-report to the tax authorities. From this setting, Bazart and Bonein implemented six treatments in a between-subjects design[102] to take into account two sources of unfairness: tax rules, and others' evasion through the provision of information on fellow citizens' average declarations. The experimental treatments were the following: (i) a benchmark treatment, in which subjects were not provided with any kind of information about the declaration of other group members; (ii) two vertical inequity treatments in which tax rates differed for fiscally identical taxpayers (being either higher or lower than the benchmark rate) but no social information was provided to subjects; (iii) a horizontal inequity treatment in which social information on the average declaration of other group members was provided;[103] and (iv) two additional treatments in which vertical and horizontal inequities co-existed.

A total of 288 subjects participated in the experiment, with 24 subjects per session who repeated the declaration game over 20 periods systematically. Nevertheless, in order to calibrate our econophysics model, we needed a homogeneous set of data in which taxpayers of the four types could coexist. Consequently, we restricted the sample to the horizontal inequity treatment only, and used the declaration choices made at each period over the 20 periods of the game by the 48 subjects pertaining to this treatment, representing a total of 960 declaration decisions. Bazart and Bonein (2014) showed that some taxpayers did change their declaration decisions in the next period, to get closer to the average reported income of other group members. This behaviour was qualified as reciprocal, in that it was conditional on what the others did. Bazart and Bonein (2014) demonstrated that both horizontal positive and negative reciprocity were at stake in the experiment, meaning that, if the other group members declared more (or less) on average, the subject would increase (or decrease) his report. This is classified here under the imitating type, with the difference that a taxpayer of the imitating type will have an invariant behaviour toward copying what the others do. For this reason, the imitating type of taxpayer in our econophysics model should adjust his behaviour to that of the other

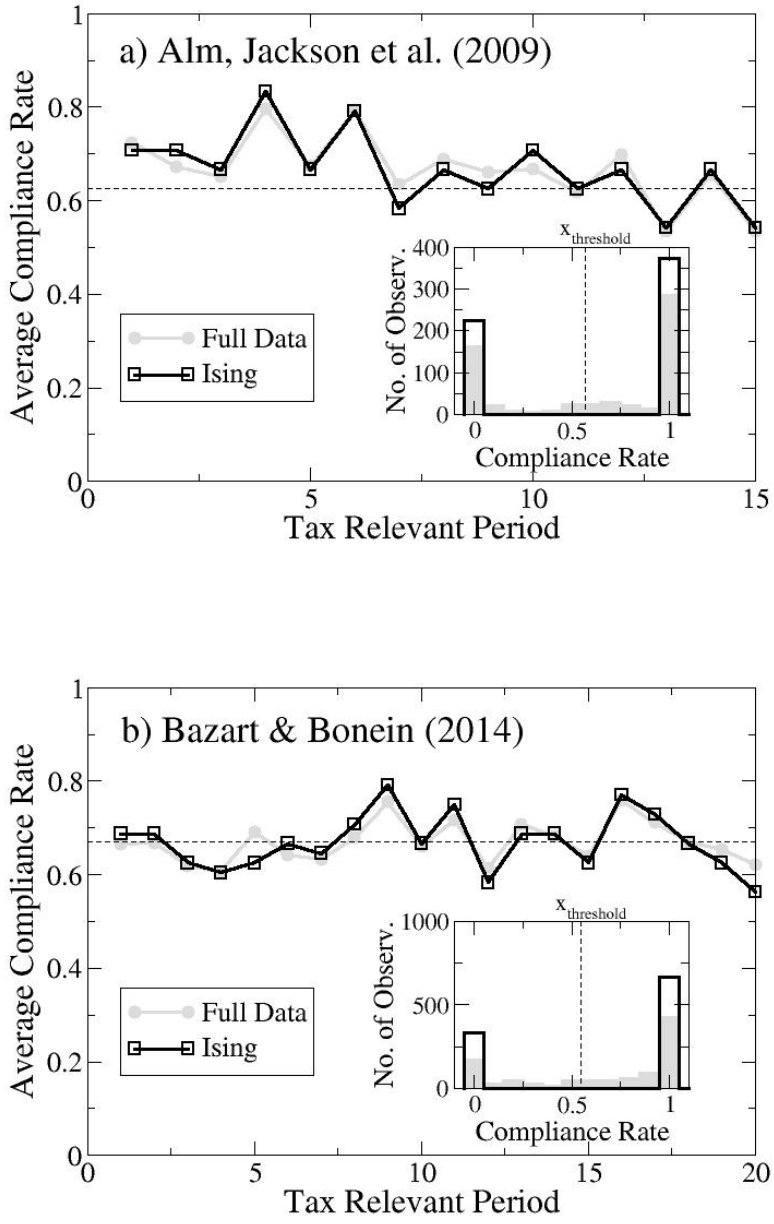
[102] For a detailed description of the design, see Bazart and Bonein (2014).
[103] Horizontal inequity results from the heterogeneity of declaration decisions in the group.

group members in all 20 periods of the game. To avoid any bias linked with the history of gains in this income declaration game, the subjects' payments corresponded with the gains of five periods randomly drawn from the 20 tax-relevant periods.

In the next section, we analyse the experimental data in the settings shown in Table 2 to extract temperature and field parameters for participants in our agent-based econophysics tax evasion model.

Calibrations

Figure 2. Main panels: Comparison between Ising data (squares) and experimental data (full points) for the average compliance rate.



Note: Panel (a) refers to Alm, Jackson, and McKee's (2009) dataset, and panel (b) is for Bazart and Bonein's (2014) dataset. The horizontal dashed line corresponds with the average over time (average compliance rate of 0.62 for Alm, Jackson, and McKee, 2009, and 0.67 for Bazart & Bonein, 2014). The insets report the frequency distribution (number of observations) of the reported income over all periods compared with the bimodal Ising distribution. The vertical dashed line marks the threshold ($x_{threshold}$ = 0.57 for Alm, Jackson, and McKee, 2009, and $x_{threshold}$ = 0.55 for Bazart and Bonein, 2014), which is used to convert real data into Ising data.

As outlined previously, at first glance it seems a severe simplification to consider 'Ising' agents that declare either zero or full income, but data from tax compliance laboratory experiments (Alm, Jackson, & McKee, 1992; Alm & McKee, 2006; Alm, Denkins, & McKee, 2009; Bloomquist, 2011a; Alm, Bloomquist, & McKee, 2015), as well as data from the IRS NRP for small business filers, support a bimodal distribution of the compliance rate which peaks at zero and full income. The same effect is observed in Alm, Jackson, and McKee's (2009) and Bazart and Bonein's (2014) experimental data (see insets to Figure 2, 600 and 960 observations, respectively), which show major peaks in the frequency of the compliance rate at 0 and 1. In Alm, Jackson, & McKee's (2009) data (for audit probability $p_a = 2/5$), this kind of behaviour is even more pronounced.

We now adopt the following procedure to transform the experimental data x_{data} to Ising data x_{ising}. A declaration $x_{data} \leq x_{threshold}$ is taken as $x_{ising} = -1$ (zero declaration), whereas for $x_{data} > x_{threshold}$ we set $x_{ising} = 1$ (full declaration). $x_{threshold}$ is chosen, such that we obtain the same average compliance rate, averaged over periods and participants (~0.62 and ~0.67 for Alm, Jackson, & McKee's, 2009 and Bazart & Bonein's, 2014 experiments, respectively) for the experimental and Ising data. This average is marked by the horizontal dashed line in the main panels of Figure 2. As a result, we obtain $x_{threshold}$ = 0.57 for Alm, Jackson, & McKee's (2009) data, and $x_{threshold}$ = 0.55 for Bazart and Bonein's (2014) data. Moreover, it can be seen from the main panels of Figure 2 that the temporal evolution of both datasets is very close, which further validates our mapping procedure.

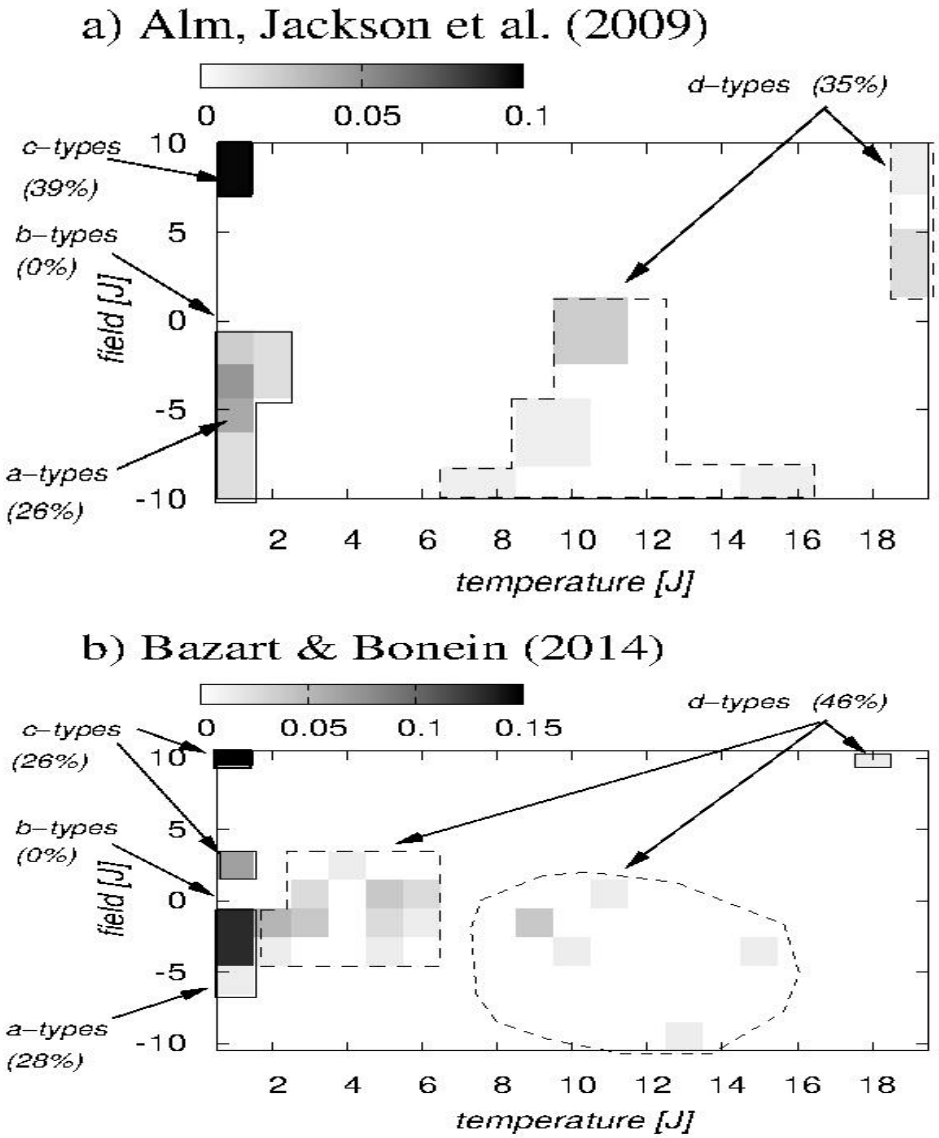
Based on the Ising dataset, we determine a local temperature T_i and magnetic field B_i parameter for each participant. For this procedure, we use Equation (2), which determines the probability $p_i = p_i(T_i, B_i)$ for the transition S_i to $-S_i$ for agent *i*. Since p_i depends on the state of neighbours, we first collect, for each participant in a given state S_i with a given neighbour configuration, the number of periods where this arrangement is the same. We then check whether or not the agent has changed her state in the next period, which allows for determination of the transition probability for a fixed neighbour configuration. Since we need two equations to determine the two variables T_i and B_i, we repeat the same procedure for another neighbour configuration. In practice, we take those neighbour configurations which occur most frequently within the time period of the experiment.

In this way, local temperature and magnetic field parameters are determined for each participant, and in Figure 3 we show the resulting distribution of parameter values for the data from Alm, Jackson, and McKee (2009) in Panel (a) and from

Bazart and Bonein (2014) in Panel (b). There are several notable points. First, there is no indication of a pure imitating b-type which, as noted earlier, is specified by $B_i = 0$ and $T_i \ll J$. Second, in both datasets, the percentage of a-types is comparable (26 per cent in Alm, Jackson, & McKee, 2009, and 28 per cent in Bazart & Bonein, 2014, respectively) and the majority of selfish a-types are of the same order as the interaction energy coming from nearest neighbours. Therefore, these types are not purely non-compliant but have a significant tendency to copy the behaviour of their social network. The same holds for the low-temperature d-types (~26 per cent) in Bazart and Bonein's (2014) data, which probability is also not purely random but also influenced by the state of nearest neighbours. On the other hand, we see from Figure 3b) that there is a second group of high-temperature d-types (~ 20 per cent) which, in all periods, make purely random decisions between compliance and non-compliance. The same holds for the 35 per cent of d-types in Alm, Jackson, and McKee's (2009) data.

Analysis of the agent distribution, shown in Figures 3a) and 3b), reveals interesting differences. This concerns, in particular, the percentage of c- and d-types, while the shares of (pure) b-types (0 per cent) and a-types (~26-28 per cent) are similar. In fact, the 35 per cent d-types and 39 per cent c-types in Alm, Jackson, and McKee's (2009) data appear as 46 per cent d-types and 26 per cent c-types in analysis of Bazart and Bonein's (2014) experiment, which needs explanation. In both experiments, participants were drawn from a pool of undergraduate students. Although there may have been differences in their sociocultural background (European versus US) and there is also a slight difference in audit probabilities (2/5 versus 1/3), this does not account for the difference of 10 per cent in the c- and d-type compositions. A rather more plausible explanation relates to the experimental design concerning the income of participants. While, in Alm, Jackson, and McKee's (2009) experimental design, participants earned income through their performance in a task based on 20 periods, the setting of Bazart and Bonein (2014) was such that individuals in each period received a constant income of 100 points and were paid for five randomly-selected periods. It is likely that income resulting from labour rather than as a "lump sum" was valued more highly because a high wage in one period did not guarantee the same wage in the next period. Therefore, individuals may have been more careful in managing their assets, which was in turn reflected in the increased moral attitude of the participants. On the other hand, participants who were always sure of receiving the same fixed wage in the next period may have been more susceptible to evading part of this income in order effectively to increase their assets. Such sporadic evading behaviour is characteristic of d-types, which may explain their larger percentage in Bazart and Bonein's (2014) experiment. It would be interesting to investigate this hypothesis in a future experimental study.

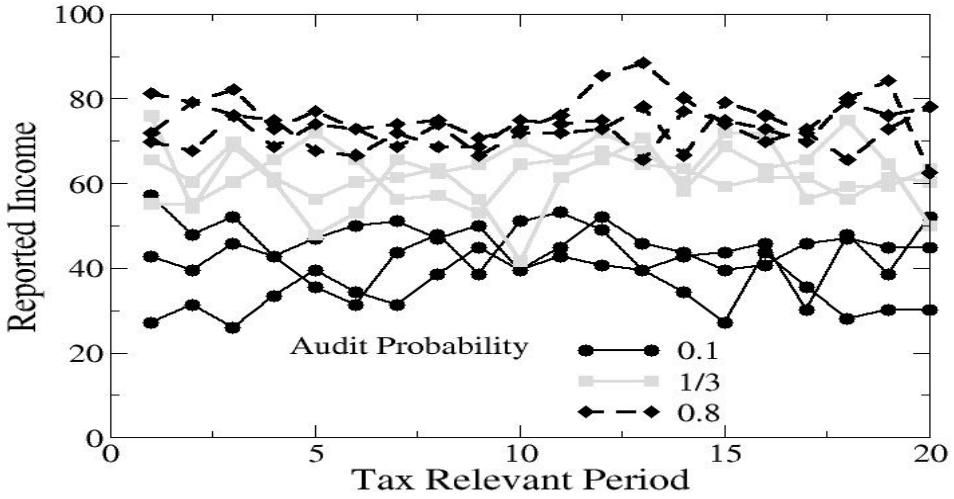
Figure 3. Parameter distribution of agent types determined from the Ising data



Note: The borderlines group the agents according to the classification of types. Panel (a) shows data from Alm, Jackson, and McKee (2009); Panel (b) shows data from Bazart and Bonein (2014).

Having characterized all participants in the tax compliance experiment by their local temperature and magnetic field parameters, we are now in a position to evaluate and predict the time-dependent reported income for different experimental settings. In Figure 4, this is exemplified for a hypothetical experiment which differs from Bazart and Bonein (2014) only in the audit probability. For each audit probability, we show three simulations which differ within the error induced by the finite group size. The interesting finding concerns the relatively small increase in reported income, from ~40 per cent to ~78 per cent, on increasing the audit probability from $p_a = 0.1$ to $p_a = 0.8$. The reason for this weak dependence on p_a can be traced back to the large fraction of d-types (~46 per cent) among the participants. Since they predominantly declare randomly, these agents are only weakly affected by an audit. Of course, this conclusion only holds when the distribution obtained in Figure 3 itself only weakly depends on audit probability.

Figure 4. Results for time-dependent reported income evaluated against the agent distribution reported in Fig. 3a and different audit probabilities



Note: For each audit probability, three simulations were realized, which differ due to finite size fluctuations.

CONCLUSIONS

In this paper, we have presented calibrations of our agent-based econophysics tax evasion model based on Pickhardt and Seibold (2014), with experimental tax compliance data taken from Alm, Jackson, and McKee (2009) and Bazart and Bonein (2014). To the best of our knowledge, this kind of calibration has never been done before in the econophysics branch of agent-based tax evasion modelling. Moreover, following the discussions in Schulz (2003), Zaklan, Westerhoff, and Stauffer (2009), Hokamp and Pickhardt (2010) and Pickhardt and Seibold (2014), we have given an economic interpretation of physical quantities in econophysics. For instance, magnetic fields reflect a moral attitude of agents, and local temperature measures the susceptibility of agents to external perturbations. According to our analysis, the pure agent types introduced in Hokamp and Pickhardt (2010) and Pickhardt and Seibold (2014) are not visible in participants in the tax compliance experiments conducted by Alm, Jackson, and McKee (2009) and Bazart and Bonein (2014). Rather, we find agent types whose behaviour is a mixture of non-compliant and imitating (a-types), compliant and imitating (c-types), and random and imitating (d-types). Only for the d-types, there also exists a pure sub-group with a large temperature parameter, so that agents act purely randomly over all periods.

Furthermore, we have been able to replicate findings frequently observed in tax compliance experiments, in particular regarding wide fluctuations in subjects' income declaration behaviour. This result is due to downsizing the population of our agent-based econophysics model (from 10^6 to fewer than 50 agents). Cline, Bloomquist, Gentile, Koehler and Marques (2013) and Bloomquist and Koehler

(2015) conduct their research in the contrary direction; that is, they build a large-scale agent-based model of tax compliance (~10^8 agents). Our findings support their notion that scale influences aggregate taxpayer behaviour in computational social simulations. Within our agent-based econophysics approach, these differences between large- and small-scale tax evasion simulations are due to (i) enhanced statistical fluctuations relating to small group sizes, and (ii) alterations in social network structures regarding small-scale experimental designs and large-scale *real world* situations.

However, the calibrations carried out in this paper are only a first step toward establishing an agent-based econophysics approach to tax evasion dynamics. In particular, it is important to analyse whether, in our approach, the agent-type distribution is dependent on the experimental setting, for example whether it depends on audit probability. In addition, it may be that agents change their behaviour over time, so that the local temperature and magnetic field parameters acquire a temporal dependence. Further research is required to allow the forecasting of tax evasion through agent-based modelling.

REFERENCES

Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics, 1,* 323-338.

Alm, J. (2010). Testing behavioral public economic theories in the laboratory. *National Tax Journal, 63*(4), 635-658.

Alm, J. (2012). Measuring, explaining, and controlling tax evasion: Lessons from theory, experiments, and field studies. *International Tax and Public Finance, 19*(1), 54-77.

Alm, J., Bloomquist, K. M., & McKee, M. (2015). On the external validity of laboratory tax compliance experiments. *Economic Inquiry, 52*(2), 1170-1186.

Alm, J., Deskins, J., & McKee, M. (2009). Do individuals comply on income not reported by their employer? *Public Finance Review, 37,* 120-141.

Alm, J., Jackson, B. R., & McKee, M. (1992). Institutional uncertainty and taxpayer compliance. *The American Economic Review, 82*(4), 1018-1026.

Alm, J., Jackson, B. R., & McKee, M. (2009). Getting the word out: Enforcement information dissemination and compliance behavior. *Journal of Public Economics, 93,* 392-402.

Alm, J., & McKee, M. (2006). Audit certainty, audit productivity, and taxpayer compliance. *National Tax Journal, 59,* 801-816.

Arsian, O., & İcan, Ö. (2013a). An agent-based analysis of tax compliance for Turkey. *Anadolu University Journal of Social Sciences, 13*(2), 143-152.

Arsian, O., & İcan, Ö. (2013b). The effects of neighborhood on tax compliance rates: Evidence from an agent-based model. *Ç. Ü. Sosyal Bilimler Enstitüsü Dergisi, 22*(1), 337-350.

Bazart, C., & Bonein, A. (2014). Reciprocal relationships in tax compliance decisions. *Journal of Economic Psychology, 40,* 83-102.

Bazart, C., & Pickhardt, M. (2011). Fighting income tax evasion with positive rewards. *Public Finance Review, 39*(1), 124-149.

Bloomquist, K. M. (2004a). *Modeling Taxpayers' Response to Compliance Improvement Alternatives*. Paper presented at the Annual Conference of the North American Association for Computational Social and Organizational Sciences, Pittsburgh, PA, 27-29 June.

Bloomquist, K. M. (2004b). *Multi-agent Based Simulation of the Deterrent Effects of Taxpayer Audits*. Paper presented at the 97th Annual Conference of the National Tax Association, Minneapolis, MN, November.

Bloomquist, K. M. (2006). A comparison of agent-based models of income tax evasion. *Social Science Computer Review, 24*(4), 411-425.

Bloomquist, K. M. (2008). Tax compliance simulation: A multi-agent based approach. In B. Edmonds, K. G. Troitzsch, & C. H. Iglesias (Eds), *Social Simulation: Technologies, Advances and New Discoveries* (pp.13-25). Hershey, PA: Information Science Reference.

Bloomquist, K. M. (2011a). Tax compliance as an evolutionary coordination game: An agent-based approach. *Public Finance Review, 39*(1), 25-49.

Bloomquist, K. M. (2011b). *Multi-agent Simulation of Taxpayer Reporting Compliance*. Invited Paper presented at the International Conference on Taxation Analysis and Research, London, December.

Bloomquist, K. M. (2012). *Agent-based Simulation of Tax Reporting Compliance*. PhD Thesis, George Mason University, Fairfax, VA.

Bloomquist, K. M., & Koehler, M. (2015). A large-scale agent-based model of tax reporting compliance. *Journal of Artificial Societies and Social Simulation, 18*(2). Retrieved from http://jasss.soc.surrey.ac.uk/18/2/20.html.

Cline, J., Bloomquist, K. M., Gentile, J. E., Koehler, M., & Marques, U. (2013). *From Thought to Action: Creating Tax Compliance Models at National Scales*. Paper presented at the International Conference on "The Shadow Economy, Tax Evasion and Governance", Münster, Germany, July.

Crokidakis, N. (2014). A three-state kinetic agent-based model to analyze tax evasion dynamics. *Physica A, 414*, 321-328.

Davis, J., Hecht, G., & Perkins, J. D. (2003), Social behaviors, enforcement and tax compliance dynamics. *The Accounting Review, 78*, 39-69.

Garrido, N., & Mittone, L. (2013). An agent based model for studying optimal tax collection policy using experimental data: The cases of Chile and Italy. *Journal of Socio-Economics, 42*, 24-30.

Hokamp, S. (2013). *Income Tax Evasion and Public Goods Provision: Theoretical Aspects and Agent-based Simulations*. PhD Thesis, Brandenburg University of Technology, Cottbus, Germany.

Hokamp, S., & Pickhardt, M. (2010). Income tax evasion in a society of heterogeneous agents: Evidence from an agent-based model. *International Economic Journal, 24*(4), 541-553.

Hokamp, S., & Seibold, G. (2014a). How much rationality tolerates the shadow economy? An agent-based econophysics approach. In B. Kamiński, & G. Koloch (Eds), *Advances in Social Simulation, ESSA 2013* (pp.119-128). Berlin: Springer.

Hokamp, S., & Seibold, G. (2014b). Tax compliance and public goods provision: An agent-based econophysics approach. *Central European Journal of Economic Modelling and Econometrics, 6*(4), 217-236.

Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik, 31*(1), 253-258.

Krauskopf, T., & Prinz, A. (2011). Methods to reanalyze tax compliance experiments: Monte-Carlo simulations and decision time analysis. *Public Finance Review, 39*(1), 168-188.

Krauth, W. (2006). *Statistical Mechanics, Algorithms and Computations.* Oxford: Oxford University Press.

Lima, F. W. S. (2010). Analysing and controlling the tax evasion dynamics via majority-vote model. *Journal of Physics: Conference Series, 246*, 1-12.

Lima, F. W. S. (2012a). Tax evasion dynamics and Zaklan model on opinion-dependent network. *International Journal of Modern Physics C: Computational Physics and Physical Computation, 23*(6), 1250047-11.

Lima, F. W. S. (2012b). Tax evasion and nonequilibrium model on Apollonian networks. *International Journal of Modern Physics C: Computational Physics and Physical Computation, 23*(11), 1250079-8.

Lima, F. W. S., & Zaklan, G. (2008). A multi-agent-based approach to tax morale. *International Journal of Modern Physics C: Computational Physics and Physical Computation, 19*(12), 1797-1808.

Llácer, T., Miguel, F. J., Noguera, J. A., & Tapia, E. (2013). An agent-based model of tax compliance: An application to the Spanish case. *Advances in Complex Systems, 16,* 1350007-33.

Macal, C. M., & North, M. J. (2005). Tutorial on agent-based modeling and simulation. In M. E. Kuhl, N. M. Steiger, F. B. Armstrong, & J. A. Joines (Eds), *Proceedings of the 37th Winter Simulation Conference* (pp.2-15), Orlando, FL, 4-7 December.

Miguel, F. J., Noguera, J. A., Llácer, T., & Tapia, E. (2012). *Exploring Tax Compliance: An Agent-based Simulation*. Paper presented at the 26th European Conference on Modelling and Simulation, Koblenz, Germany, May/June.

Mittone, L., & Patelli, P. (2000). Imitative behavior in tax evasion. In F. Luna, & B. Stefansson (Eds), *Economic Simulations in Swarm: Agent-based Modelling and Object Oriented Programming* (pp.133-158). Dordrecht: Kluwer Academic.

Noguera, J. A., Llácer, T., Miguel, F. J., & Tapia, E. (2014). Tax compliance, rational choice, and social influence: An agent-based model. *Revue Française de Sociologie, 55*(4), 765-804.

Nordblom, K., & Žamac, J. (2012). Endogenous norm formation over the life cycle: The case of tax morale. *Economic Analysis and Policy, 42*(2), 153-170.

Pickhardt, M., & Prinz, A. (2014). Behavioral dynamics of tax evasion: A survey. *Journal of Economic Psychology, 40*, 1-19.

Pickhardt, M., & Seibold, G. (2014). Income tax evasion dynamics: Evidence from an agent-based econophysics model. *Journal of Economic Psychology, 40*, 147-160.

Rand, W., & Trust, R. T. (2011). Agent-based modeling in marketing: Guidelines for rigor. *International Journal of Research in Marketing, 28,* 181-193.

Schinckus, C. (2013). Between complexity of modelling and modelling of complexity: An essay on econophysics. *Physica A, 392*(13), 3654-3665.

Schulz, M. (2003). *Statistical Physics and Economics: Concepts, Tools, and Applications*. New York: Springer.

Seibold, G., & Pickhardt, M. (2013). Lapse of time effects on tax evasion in an agent-based econophysics model. *Physica A, 392*(9), 2079-2087.

Sornette, D. (2014). Physics and financial economics (1776-2014): Puzzles, Ising and agent-based models. *Reports on Progress in Physics, 77*, 062001-1-28.

Srinivasan, T. N. (1973). Tax evasion: A model. *Journal of Public Economics, 2*, 339-346.

Stauffer, D. (2013). A biased review of sociophysics. *Journal of Statistical Physics, 151*, 9-20.

Train, K. (2009). *Discrete Choice Methods with Simulation.* Cambridge: Cambridge University Press.

Yitzhaki, S. (1974). A note on income tax evasion: A theoretical analysis. *Journal of Public Economics, 3*, 201-202.

Zaklan, G., Lima, F. W. S., & Westerhoff, F. (2008). Controlling tax evasion fluctuations. *Physica A: Statistical Mechanics and its Applications, 387*(23), 5857-5861.

Zaklan, G., Westerhoff, F., & Stauffer, D. (2009). Analysing tax evasion dynamics via the Ising model. *Journal of Economic Interaction and Coordination, 4*, 1-14.

The Taxation of Micro and Small Businesses in Transition Economies: Country Experience of the Introduction of Special Tax Regimes

Michael Engelschalk[104] *& Jan Loeprick*[105]

Abstract

This paper analyzes the design of simplified small business tax regimes in Eastern Europe and Central Asia and the impact of such regimes on small business tax compliance. While there are many options for tax simplification, the general trend in the region is to offer small businesses the option to be taxed based on their turnover rather than net income. This study finds that many of the regimes in place are not well targeted, and neither take into account fairness considerations nor facilitate business growth and migration to the standard tax regime. While revenue generation is not a main objective of such regimes, the extremely low revenue performance and the risk of system abuse by larger businesses should be issues of concern. More attention should therefore be devoted to improving the design of simplified regimes and monitoring their application. This will require, in particular, a more profound analysis of the economic situation and the tax compliance challenges facing the small business segment, and increased efforts to improve the quality of bookkeeping.

JEL codes: H25, H26, O17

Keywords: Presumptive taxation; business formalization and growth; tax compliance; compliance costs.

INTRODUCTION

Business taxation is at the heart of the relationship between the state and its economic constituents. The transition toward market principles in the Eastern Europe and Central Asia (ECA) region has thus required a fundamental change in the principles underlying public revenue collection for a growing number of private enterprises: a move away from reliance on transfers, typically predetermined by state owned enterprises (SOEs), toward the assessment of actual taxable income.

As part of this major change in revenue administration and wider privatization and deregulation efforts, many transition countries established special programmes to administer and support the growth and competitiveness of micro, small and medium enterprises (MSMEs). The development of small business activity during the transition resulted in major administrative challenges and a range of policy experiments to

[104] Senior public sector specialist, World Bank Group, Governance Global Practice (mengelschalk@worldbank.org).
[105] Senior economist, World Bank Group, Governance Global Practice (jloeprick@worldbank.org).

address these, not least regarding their tax treatment. Facing an ever-growing number of business "clients", the introduction of various simplified taxation schemes was partly an attempt by tax policy makers to alleviate the administrative workload. In practice, however, audit coverage remained relatively intense, given the common starting point of regular audits of all businesses before the transition (Engelschalk, 2004). At the same time, starting with De Soto's influential work (1989), the extensive informality of small firms and individual entrepreneurs has attracted increasing attention as a challenge for transition economies (Schneider & Enste, 2000), with simplified tax policy and administrative requirements commonly seen as one of the main policy levers.

In light of these ambitious objectives, experience over the last 20 years casts some doubt on the effectiveness of simplified and preferential tax treatment in reducing compliance costs and burdens for tax administrations, or in improving formal business creation and small enterprise growth. In some instances, widespread non-compliance and under-reporting linked to simplified taxation may become a major constraint on investment activities, as unfair competition from businesses availing themselves of tax avoidance schemes increases economic pressure on formal businesses in the standard tax regime and reduces their competitiveness (World Bank, 2011).

Research on a number of tax challenges for transition economies, in particular regarding the use of presumptive taxation regimes and the control of rampant corruption, has been scarce (Holmes, 2002), aside from general guidance on MSME taxation (ITD, 2007; IFC, 2007) and select efforts to summarize country practices (Engelschalk, 2004). Little evidence is on offer for policy makers interested in how to proceed in these areas of second-, or rather third-best, policy and administrative solutions, which are prevalent given persisting capacity and resource constraints on both taxpayers and tax administrations.

This paper aims to contribute to filling this gap based on documented country experience. We focus on the tax treatment of micro and small businesses,[106] and provide an overview of policy trends and a general discussion of country experiences in MSME tax policy in the region. This country practice then informs a summary of

[106] Most countries in the region have a general definition of what constitutes a micro and a small business, either in a special SME law or in commercial laws, and all countries define MSMEs for statistical purposes. These definitions generally refer to several parameters. In the Russian Federation, for example, a business is considered to be small if annual turnover is not more than 11.2 million US$ and the number of employees does not exceed 100 (Law 209-FZ). In Croatia, the national accounting law sets small business thresholds at asset values below 4.8 million US$, annual revenues below 9.5 million US$, and average number of employees during the business year not more than 50. However, such definitions are generally not relevant for taxation purposes. Tax laws include special micro and small business thresholds, based on the revenue potential of the segment and its compliance capacity. These definitions are generally turnover-based and are discussed in more detail later in the paper.

lessons learned and policy recommendations, which are presented in the second part of the paper.

Regional overview: Development and issues regarding the tax treatment of MSMEs

Income/profit taxes and compliance costs

In the ECA region, presumptive tax instruments were typically introduced in the late 1990s or early 2000s, with the objective of promoting private-sector development and facilitating compliance management in an environment characterized by low tax administration capacity and a rapidly growing number of private small business operators. Prioritizing the effective compliance management of larger businesses to ensure sufficient revenue mobilization required tools to minimize the administrative efforts of smaller entities. The goal was to encourage voluntary compliance of small businesses, while allowing for a simple examination of low-revenue tax returns.

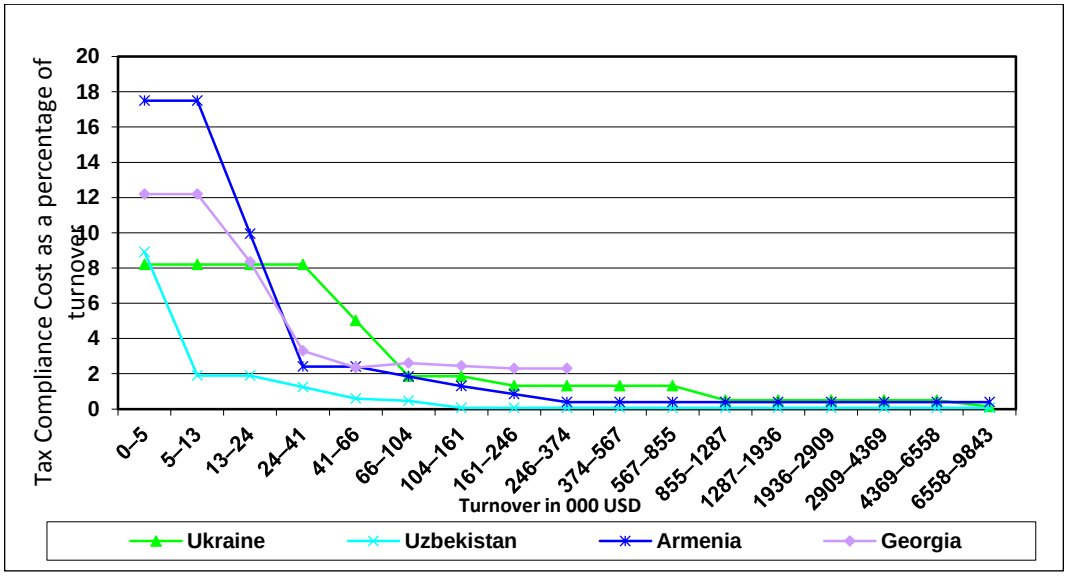
Moreover, a number of compliance burden and compliance cost studies has highlighted regressive features of tax compliance costs and stressed the need for developing simplified systems of taxation. Klun and Blasic (2005) provide estimates for Slovenia and Croatia and, similarly, survey-based analysis by the World Bank in Ukraine (2009), Uzbekistan (2008), Armenia (2010) and Georgia (2011) supports earlier findings on compliance cost assessment in the OECD, depicting a highly regressive burden. Given the high fixed-cost component of tax compliance, the general trend identified in these surveys is hardly surprising: the smaller the business, the higher the tax compliance cost it faces as a share of its turnover.

It is notable that, even for businesses operating at more than $100,000 in turnover, measured compliance costs may still surpass three per cent of their turnover level. The reasons for such high compliance burdens vary, including complicated reporting procedures and time spent on inspection visits and/or audits, a frequent challenge for MSMEs in the region (Engelschalk & Loeprick, 2011). For example, Kireeva and Rudy highlight that, in Belarus, to comply with the general tax regime, a small business with up to 50 employees on its books has to employ an average of two accountants. They estimate that monthly costs incurred by the SME segment on tax compliance may exceed $3.5 million (National Report for Belarus, in Lang, Obermair, Schuch, Staringer, & Weninger, 2008, pp. 91-110).

Compliance costs may be burdensome, irrespective of legal status, and participation in the presumptive regime may be a tool to reduce these and thereby increase the competitiveness of small corporations. Moreover, small corporations do not necessarily have better in-house accounting capacity than non-incorporated businesses. On the other hand, however, different accounting obligations may already require more

comprehensive records for corporations, and the risks of system abuse and downward migration increase markedly when corporations are eligible for simplified income tax treatment.

Figure 1. Regressive tax compliance costs for micro and small businesses in the region



Source: IFC Tax Compliance Cost Surveys, 2007-11.

Attempts made in a number of countries to abolish presumptive tax regimes for small businesses have generally been unsustainable. For example, in Georgia, a patent system for small businesses was in place until 2005, when a new tax code was introduced. However, taxation of small businesses based on net income resulted in high compliance costs, and in 2011 Georgia adopted a new simplified tax regime. Micro businesses with a turnover below GEL 30,000 ($18,100) are exempt from income taxation, while small businesses with a turnover below GEL 100,000 pay a presumptive tax based on turnover. Romania operated a micro-enterprise tax (MET) regime with a three per cent rate on turnover until January 2010, when the regime was abolished and small taxpayers were moved to the general tax regime. A year later, the system was reintroduced and, starting from January 2013, presumptive taxation became mandatory even for incorporated small businesses. The system was used by 92,000 taxpayers prior to its abolition, representing about 20 per cent of eligible small businesses, and around 60,000 of these businesses immediately moved back to presumptive taxation after its reintroduction. Armenia had a turnover tax for businesses with turnover below AMD 30 million ($71,700). This was abolished in 2008 as part of a major tax reform exercise, which also increased the VAT threshold to AMD 58.25 million ($180,000). Only the patent regime for micro businesses and a presumptive small business tax for a few selected activities, such as barber shops, remained in place. The repeal of the turnover

tax regime was motivated partly by widespread abuse of the regime by larger businesses. However, it resulted in a significant additional compliance burden for many small businesses, and therefore presumptive taxation based on turnover was reintroduced in 2013.

Basic system design

Most regimes targeting micro, small and medium-sized businesses in the region have changed fundamentally since their first introduction, and many regimes are modified on a regular basis. While, in the 1990s, simple fixed-tax or patent regimes were also widely used for the small business segment, today, such regimes are largely limited to micro businesses, and a turnover-based approach has become the standard method for taxing small businesses (see Table 1).

Table 1. Approach to simplified small business taxation, 2000 and 2014

Country	Tax regime 2000[107]	Tax regime 2014
Albania	Turnover	Net income
Armenia	Lump sum	Turnover
Azerbaijan	Turnover	Turnover
Belarus	Lump sum	Turnover
Bosnia	None	Turnover
Bulgaria	Lump sum	Lump sum
Croatia	Lump sum	Net income
Czech Republic	Lump sum	Standard deduction from gross income
Estonia	Net income tax	Net income tax
Georgia	None	Turnover
Hungary	None	Standard deduction from gross income
Kazakhstan	Lump sum or turnover	Turnover
Kosovo	Turnover	Turnover
Kyrgyz Republic	Turnover	Turnover
Latvia	None	Turnover
Lithuania	Presumptive tax	Lump sum
Macedonia	Lump sum	Turnover
Moldova	Lump sum	Turnover
Poland	Lump sum	Turnover
Romania	Turnover	Turnover
Russia	Turnover	Turnover
Serbia	None	Turnover
Slovak Republic	Lump sum	Standard deduction from gross income
Slovenia	None	Standard deduction from gross income
Ukraine	Turnover	Turnover
Uzbekistan	Turnover	Turnover

Source: Authors, based on Mitra and Stern, 2002 and World Bank Group Country Reports.

Experimentation with the tax treatment has been common in many countries. Kazakhstan, for example, first introduced simplified taxation for micro and small

[107] Based on Mitra & Stern, 2002

businesses in 1995. The system has been modified several times since, including a move from a fixed tax to a turnover-based tax and the replacement of a progressive presumptive tax with a single rate on turnover.

Table 2. Presumptive tax regimes for micro and small businesses in Kazakhstan

<table>
<tr><th>—1995</th><th>1995</th><th>2001</th><th>2002</th><th>2004</th><th>2007—</th></tr>
<tr><td rowspan="4">No presumptive regime</td><td colspan="5">Patent regime</td></tr>
<tr><td rowspan="3">Introduction of the patent regime with tax rates differentiated by type of profession (c.150 different professions)</td><td>Limitation of the patent regime to individual entrepreneurs</td><td>Replacement of fixed patent with flat 3% rate on turnover</td><td>Increase of the turnover threshold (aligning it with minimum monthly wages)</td><td>Flat tax rate lowered to 2%</td></tr>
<tr><td colspan="4">Simplified declaration regime</td></tr>
<tr><td>Introduction of simplified declaration regime with progressive tax scale (4-11% of turnover for physical entities; 5-13% for corporations</td><td>Rate reduction to 3-7% for physical entities and 4-9% for corporations</td><td>Further reduction in tax liabilities (progression limited to 3-5% for physical entities and 3-7% for corporations)</td><td>Introduction of a flat 3% rate on turnover</td></tr>
</table>

Source: World Bank Group, 2010.

In addition to the move toward a turnover-based calculation of presumptive tax liability, it was increasingly recognized that the MSE segment of the taxpayer population actually consists of two different taxpayer groups: micro and small businesses. Further segmentation to differentiate between the small and micro business categories constitutes the second major system reform trend (see below for a definition of micro businesses). Frequently, segmentation was combined with an effort to improve local revenue mobilization, and micro business tax revenues were allocated to local budgets.

Flat income taxation and simplified small business regimes

The ECA region is not only a region with widespread use of presumptive tax regimes for small businesses; it is also a region in which many countries have introduced flat income tax regimes, an approach pioneered by Estonia in 1994. Following Estonia and its Baltic neighbors, the Russian flat tax reform in 2001 attracted global attention as a result of subsequent improvements in revenue collection. This triggered a wave of

similar reform efforts throughout the region (see Table 3). The general objective was to promote economic growth through creation of a business- and investment-friendly environment for individuals and companies, as well as to achieve a high degree of tax fairness (see Brook & Leibfritz. 2005), to simplify administration and compliance, and to introduce greater tax transparency.

Table 3. Overview of spread of flat tax reforms in ECA

Country	**Year flat tax introduced**	**Rate (%)**	**Impact on small business regimes**
Albania	2007 (abolished 2014)	10	Presumptive regime had been transferred to local governments and remained in place
Bosnia & Herzegovina	2009	10	None
Bulgaria	2007	10	Presumptive tax transferred to local governments in 2008
Czech Republic	2008	15	Lump-sum deduction scheme already in place before flat tax was introduced and maintained
Estonia	1994	21[108]	No special MSME tax regime developed
Georgia	2005	12	Presumptive regime abolished with new tax code, but later re-introduced
Kyrgyz Republic	2006	10	None
Latvia	1997	25	Micro-enterprise tax introduced in 2010
Lithuania	1994	33	N/A
Macedonia	2007	10	N/A
Montenegro	2009	9	N/A
Romania	2005	16	Turnover tax regime introduced in 2004, remained in place
Russia	2001	13	Simplified tax regime introduced in 2003
Slovak Republic	2004 (abolished 2013)	19	Presumptive regime abolished with flat tax introduction
Ukraine	2004	13	Simplified regime introduced in 1998, remained unchanged

The move toward flat income taxation typically affected a broad range of related taxation areas,[109] although details of the reform programmes differed, as summarized by Keen, Kim, and Varsano (2008). The revenue impact varied depending on the degree and effectiveness of accompanying measures to increase the tax base, economic growth at the time of the reform, and complementing enhancements of the tax administration (World Bank, 2007). Generally, flat tax regimes, which tended to provide important alleviations of the tax burden in the upper income brackets, reduced personal income tax revenues – Latvia, Lithuania and Russia being exceptions – and

[108] The rate in Estonia was initially set at 26 per cent and subsequently lowered.

[109] Spanning the treatment of corporate and capital income, reforms of indirect taxation and social contributions, and solutions chosen as measures to protect low-income groups.

triggered heavier reliance on indirect instruments such as VAT and excise (Keen et al., 2008).

In the majority of countries, the introduction of a flat tax regime was unconnected with the operation of a presumptive small business tax regime and had no impact on presumptive regime design and operation. Remarkable exceptions are the Slovak Republic and Georgia: in both cases, the introduction of a flat income tax was combined with a broader simplification of the tax regime. In the Slovak Republic, the 2004 tax reform process aimed to eliminate a large number of exemptions and special regimes, and introduce a consistent and comprehensive approach to direct taxation. As part of this process, the small business presumptive regime was replaced with a standard cost deduction ratio for the self-employed. This change in the small business taxation approach seems to have had a positive impact on voluntary tax compliance: the number of income tax returns submitted by the self-employed increased by 14.6 per cent in the first year of flat tax implementation.[110]

A similar approach was taken in Georgia, where the move to a flat income tax was combined with a comprehensive and successful tax simplification approach. The 2005 tax reform reduced the number of taxes from 22 to seven, the number of required visits to the tax office dropped sharply, and the estimated tax compliance rate increased from 35 per cent to around 80 per cent. Introduction of the flat tax was seen as an appropriate occasion to abolish the dysfunctional patent regime that had been in place since 1998. Unlike the Slovak Republic, no further simplification measures were foreseen for MSMEs, which were expected to comply with the standard income tax regime. However, following the far-reaching simplification of the general tax system, small businesses still experienced an increase in compliance requirements. While taxpayer perceptions improved dramatically among large businesses, an increasing number of small businesses identified tax administration as a key barrier to doing business following the reform. This is one factor that explains the decision to reintroduce a presumptive tax regime in 2010.

Another special case is Estonia, which introduced a simple flat tax regime early in the transition process, before a separate presumptive small business tax regime had been developed. The Estonian regime does not include any special rules or simplifications for small businesses. In an environment with a relatively highly educated and IT-literate small business community (more than 97 per cent of corporate tax returns and 93 per cent of personal income tax returns are filed electronically) and the absence of unofficial costs relating to taxation (Dickinson, 2012), a simple cash-based general taxation system has proved sufficient to support small business tax compliance.

[110] The Slovak reform combined a number of related measures, including increased labour market flexibility and a range of indirect tax reforms.

The Georgian example demonstrates that even a successful flat tax introduction, combined with a comprehensive tax system simplification, does not automatically guarantee that special simplification rules for small businesses are no longer needed. In particular, the requirement to calculate and document business expenses and the risk of disputes about the deductibility of such expenses may be considered an additional burden of flat tax regimes compared with presumptive tax regimes. Experience in other ECA countries, such as Bulgaria, has shown that a move to a flat tax does not substantially reduce the complexity of filing and documentation requirements, making compliance with the flat tax regime still burdensome for small business operators. This is even more of a challenge when the overall tax simplification measures combined with the flat tax introduction do not go far enough. The flat tax introduction in the Russian Federation, for example, was part of an exercise to introduce a new tax code. The first part of the code became effective in 1999 and fundamentally reformed the system of tax administration, while the second part, dealing with specific taxes, was approved in 2000 and became effective from 2001. The centrepiece of the Russian reform – a single marginal personal income tax rate of 13 per cent – was followed by an impressive increase in real personal income tax revenues of about 26 per cent in the first year after its introduction. However, using micro-level data, Ivanova, Keen, and Klemm (2005) provide cautionary insights, suggesting that attribution of revenue performance to personal income tax reform alone is questionable.[111] Moreover, the new Russian tax regime still consisted of around 40 different taxes, and small businesses remained confronted with an average of 9.56 types of tax (Shetinin, Zamulin, Zhuravskaya, & Yakovlev, 2005). Despite the flat tax introduction, the move to a more simplified tax regime for small businesses thus remained a valid concern, which was addressed with the introduction of the simplified tax system (STS) in 2003. A similar development occurred in Ukraine.

Pressure to introduce or maintain presumptive tax regimes with a low effective tax burden may also build up in the case of a flat tax regime that uses a high tax rate. Flat tax reforms in the ECA region did not necessarily lower the average and marginal tax rates for small businesses. Some people saw no change in their marginal tax rates, since many governments selected the marginal rate from one of the tax brackets used previously. This happened, for instance, in Lithuania, Latvia and Georgia.[112] The rates selected in the early flat tax reforms in the Baltics corresponded either to the highest marginal rate before the reform (Lithuania, Latvia), or to the middle of pre-reform rates (Estonia). With rates that remained high in comparison with those used under the

[111] Importantly, not only personal income tax, but revenue from all major sources increased in 2001 in Russia, suggesting broader drivers contributing to the observed performance. Analysis by Ivanova et al. (2005) underscores the importance of improved compliance, which may be linked to both the policy and administrative measures of the reform.

[112] However, individuals whose marginal tax rates did not change may have experienced a change in their average tax rates.

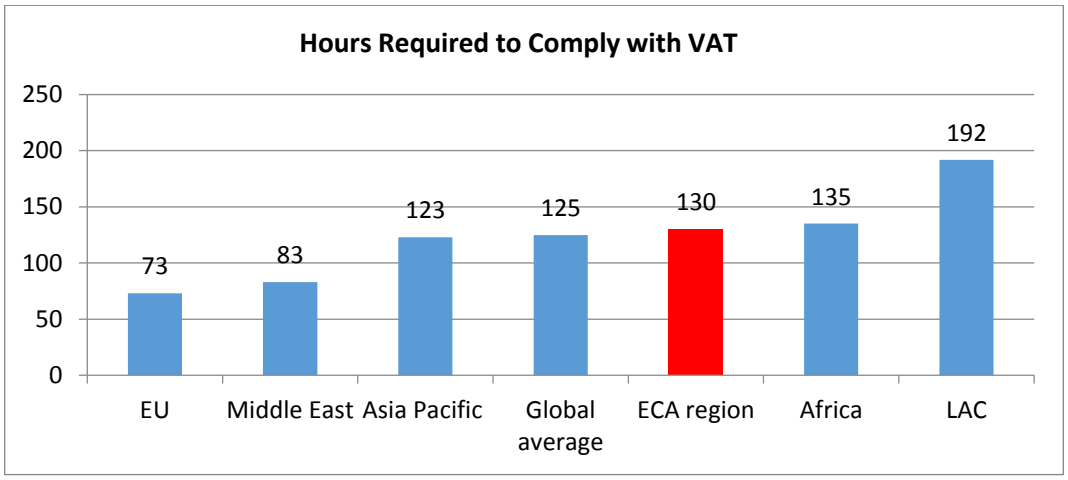
preceding income tax regime brackets (Easterbrook, 2008), small business lobby groups continued to have grounds for requesting preferential tax treatment.

Small businesses and the VAT system

In theory, strong arguments can be made in favour of including small businesses in the VAT net. As a tax on consumption, the VAT chain would ideally stretch from the point of production (or import) to the point of sale to the final consumer, thereby including the retail sector and the provision of services to private consumers. Indeed, when VAT was introduced in the ECA region in the 1990s, a number of VAT laws included no or a very low VAT threshold. For example, Russian VAT started with a very low threshold of 100,000 Russian Rubles (Rub) in 1992 (quickly increased to Rub 500,000), Hungary operated a threshold of $9,000, and Romania operated a threshold of $6,000 in the early 1990s (Jack, 1996). The VAT systems in Belarus and Uzbekistan still operate without a threshold for incorporated businesses. From the business perspective, an advantage for small businesses of being part of the VAT net might be to facilitate interaction with VAT-registered businesses, assuming that a VAT-registered larger business will prefer to order goods and services from clients who can issue VAT invoices. However, this might not always be the case; not all VAT-registered businesses are equally interested in formally deducting input VAT, and lowering the sales price of goods or services rendered may be a preferred alternative to issuing a VAT invoice. In addition, as highlighted by Bird and Gendron (2007), in some countries, vibrant markets have been established in the trading of VAT invoices. Nevertheless, good practice suggests providing small businesses with an option to register voluntarily for VAT, even if their turnover is below the registration threshold, if they can prove that they are capable and willing to comply with VAT requirements (in particular, keeping the necessary books and records, issuing VAT invoices and operating cash registers).

In practice, there are several strong arguments against an approach of extending the VAT net to the small business segment. IMF experience has shown that setting too low a threshold for VAT may significantly compromise the political and administrative feasibility of a VAT regime (Ebrill, Keen, & Summers, 2001). Furthermore, World Bank Group/IFC tax compliance cost surveys show that joining a VAT regime substantially increases tax compliance costs for small businesses. Feedback received from small businesses in European Union member countries demonstrates that small businesses consider VAT legislation to be one of the 10 most burdensome EU laws. The compliance challenge is thus increased when VAT systems require extensive documentation, where taxpayers are subject to frequent VAT audits, or when filing procedures have not been streamlined and simplified. This is the case in a number of ECA countries, as comparative analysis indicates that the time required for VAT compliance in the region substantially exceeds compliance time in EU countries.

Figure 2. VAT compliance time across regions



Source: PwC (2010).

A more detailed compliance cost analysis for VAT was conducted in Slovenia in 2001, two years after VAT had been introduced with a relatively low threshold of SIT 5 million ($22,700), demonstrating the high compliance burden for small businesses.

The risk to small businesses of complying with VAT requirements increases further in an environment of low administrative efficiency. Non-payment of VAT refunds or delays in the processing of refund requests may severely affect the liquidity of a business. Liquidity problems may also arise in the case of an accrual-based VAT system, when the VAT due has to be transferred to the Treasury before the business receives payment for goods or services from its customers. At the same time, the revenue benefits of including small businesses in the VAT net are minimal, as their contribution to total VAT collection is generally below 10 per cent.

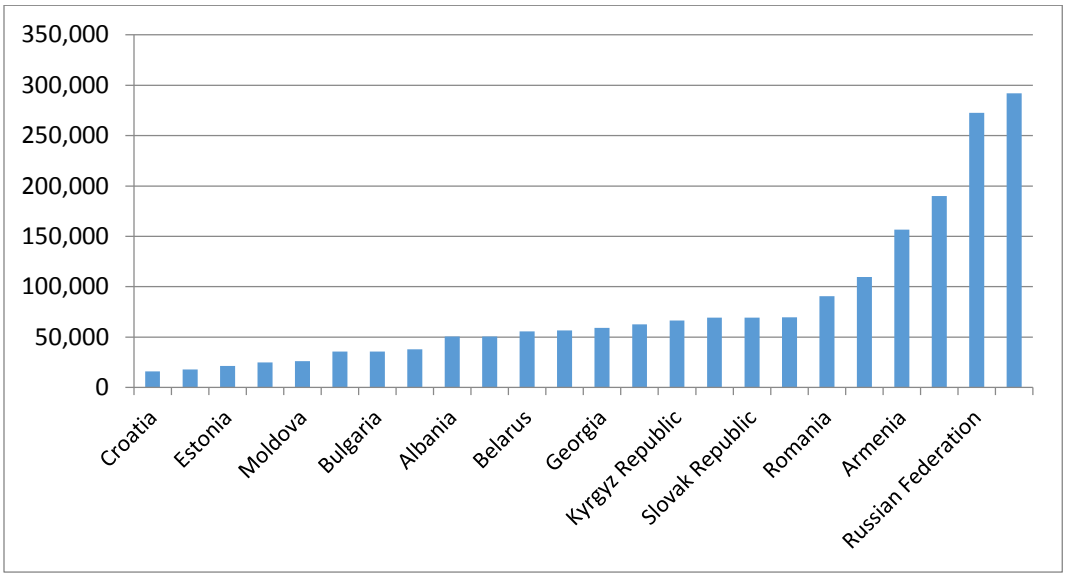
Table 4. VAT compliance costs as a share of business turnover in Slovenia

Taxpayer size(turnover)	Compliance costs as percentage of turnover
Up to SIT 100 million	3.73%
SIT 100 million–1 billion	0.73%
Above SIT 1 billion	0.08%

Source: Klun, 2003.

A reasonably high VAT registration threshold is the main tool for protecting small businesses from problems and costs relating to VAT compliance. Country practice in the region varies considerably here, and a considerable number of countries apply a rather low threshold of less than $50,000 turnover, which also risks forcing many small businesses to join the VAT net.

Figure 3. Mandatory registration thresholds for VAT in ECA (in USD)



Source: USAID (2009) – Collecting Taxes 2012-13.

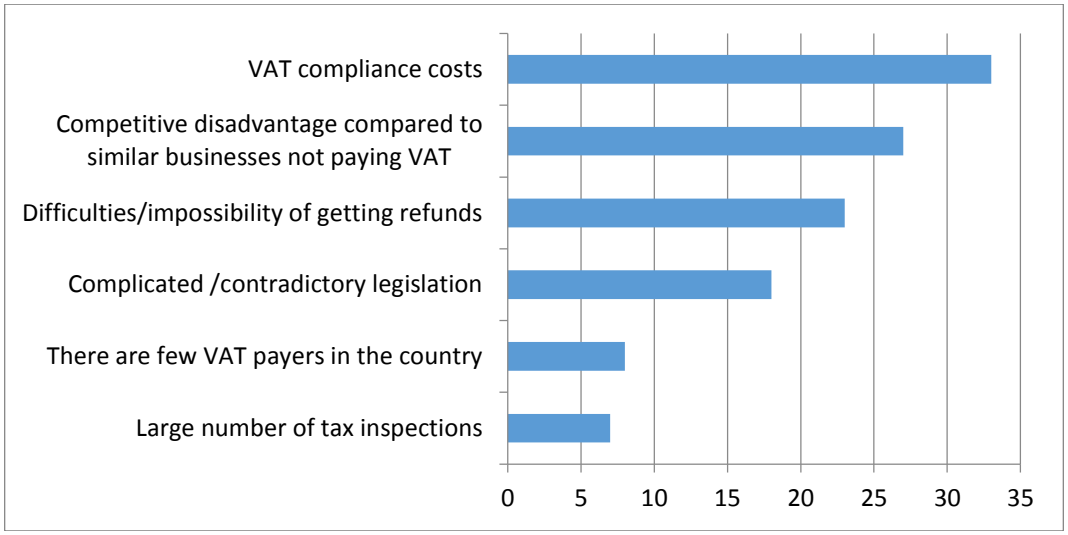
Several FSU countries (Russia, Ukraine and Belarus) have taken a different approach by integrating the VAT liability into the presumptive single tax regime. In these regimes, part of the single tax payment is considered to cover the VAT liability of the business. However, the benefits of including VAT in the single tax are questionable. From a tax administration point of view, the VAT net is not really broadened, and additional data to check the VAT compliance of larger businesses are not generated. For businesses, the tax burden is increased with the application of a higher presumptive turnover tax rate, but the business cannot deduct input VAT, nor issue VAT invoices. Consequently, the competitive position of single taxpayers supplying VAT-registered businesses does not improve. Therefore, while a single tax approach may have many benefits in general, the inclusion of VAT in the scope of tax does not generally seem sensible. The far more appropriate approach, from both the VAT revenue generation and the compliance/administration cost points of view, would be to set the VAT threshold sufficiently high to exclude small businesses from the VAT net, combined with the possibility of registering voluntarily for VAT.[113]

While the introduction of a reasonably high threshold is a convenient solution for protecting small businesses from a high VAT compliance burden, the challenges of reducing obstacles to small business growth and facilitating their transition into the VAT system remains, irrespective of the threshold level. Small businesses are likely to be reluctant to migrate into the VAT regime, even in the case of a reasonably high VAT threshold. In Uzbekistan, for example, 58 per cent of taxpayers who registered for VAT

[113] For a definition of the VAT threshold, see Keen & Mintz (2004).

saw no advantage to their business in being a VAT taxpayer, and business operators cited many disadvantages relating to joining the VAT net. Only 14 per cent of businesses saw no disadvantages to VAT registration.

Figure 4. Business perceptions of disadvantages of being VAT registered in Uzbekistan



Source: IFC (2010).

Facilitating compliance for smaller businesses in the VAT net should thus be an important part of a small business growth facilitation strategy. For many CEE countries, the EU VAT rules provide an orientation for such VAT simplification. Article 281 of the EU VAT Directive allows EU member countries to apply simplified procedures, such as flat rate schemes, for charging and collecting VAT from smaller VAT-registered businesses. The most widespread simplification measure in CEE countries is the use of VAT cash accounting schemes. These schemes allow a business to postpone its VAT payment to the date it receives payment for goods supplied and services rendered. Such schemes are in place for businesses with annual turnovers below the following thresholds for these countries: Estonia €208,646 ($236,000), Slovenia €400,000 ($453,000), Latvia (for small businesses) €100,000 ($114,000), Romania RON 2,250,000 ($682,600), Serbia (since 2013) SRD 50 million ($598,000), and Bulgaria (since 2014) €500,000 ($566,000).

Another measure with a major impact on compliance and administrative costs is a reduction in the VAT filing frequency. Analysis in the EU suggests that, for a micro business, the costs of filing monthly VAT returns amount to more than €100 (around $140) per return (European Commission, 2013). In Hungary, submitting three monthly returns rather than one quarterly return increases filing costs by 35 per cent (PwC, 2013). A move from monthly to quarterly filing would therefore reduce the annual filing costs for a micro business by $430. Quarterly filing for small businesses has now

become a widespread compliance facilitation approach, although some countries, such as Bulgaria and Estonia, require all VAT-registered businesses to file monthly VAT returns.

Table 5 shows that some countries have reduced the filing frequency still further for very small VAT payers. The most extensive approach has been taken by Hungary, offering annual VAT filing if the annual consolidated sum of the difference between the total tax payable in the second year preceding the relevant year and the tax deductible during the same year is less than Ft 250,000 ($1,100). Latvia and Lithuania offer bi-annual filing for small VAT payers.[114] Among non-EU ECA countries, several countries, such as Georgia and Russia, have simplified VAT compliance beyond the micro and small business segment with the introduction of quarterly VAT filing as a standard rule for all businesses.

While, in many countries, a reluctance by small business operators to join the VAT system can be observed, in some situations the opposite may occur. In Romania, analysis conducted in 2010 showed that almost 380,000 small businesses with a turnover below the VAT registration threshold were voluntarily VAT registered. This meant that more than 60 per cent of the VAT net comprised small businesses that, according to system design, should have remained outside the VAT system, complicating VAT administration for the tax offices and contributing only 1.3 per cent to total VAT revenues. Closer analysis is required in such situations to understand the dynamics that force small businesses into the VAT system and increase their compliance costs. A more drastic approach is to exclude micro businesses from voluntary VAT registration. This approach was tried for some time in Serbia, with the operation of a threshold for mandatory VAT registration at the level of SRD 4 million ($47,800) and a threshold for voluntary VAT registration of SRD 2 million ($23,900). The reform of the VAT law in 2012 abolished the threshold for voluntary registration, as it resulted in an obligation for registered businesses to deregister when the business turnover dropped below the SRD 2 million threshold. At the same time, the threshold for mandatory registration was increased to SRD 8 million ($95,600).

[114] In Latvia, for taxpayers with a previous year turnover below LVL (Latvian lats) 10,000 or $19,400; in Lithuania, for businesses with previous year turnover below LTL (Lithuanian centai) 200,000 or $79,300.

Table 5. Annual VAT filing frequency and number of VAT returns filed in CEE EU member states

Country	Large	Medium	Small	Micro[115]		
Bulgaria	12	12	12	12		
	429	2,361	13,953	197,917		
Czech Republic	12	12	12	12	4	
	1,006	5,531	32,681	65,472	398,093	
Estonia	12	12	12	12		
	143	785	4,640	65,818		
Hungary	12	12	12	12	4	1
	1,101	6,055	35,778	84,428	280,524	142,541
Latvia	12	12	12	12	4	2
	174	959	5,669	34,161	20,107	26,151
Lithuania	12	12	12	12		2
	148	815	4,815	52,298		16,000
Poland	12	12	12	12	4	
	3,200	17,600	104,000	1,255,200	220,000	
Romania	12	12	12	12	4	
	1,136	6,250	36,932	141,334	382,537	
Slovak Republic	12	12	12	12	4	
	393	2,162	12,774	43,733	137,456	
Slovenia	12	12	12	12	4	
	206	1,133	6,694	31,703	63,248	

Source: European Commission (2013).

Small businesses and social security systems

Social taxes constitute an important component of the tax systems of many ECA countries, and the need for compliance may be an obstacle to small business operators in formalizing and legalizing labour.

For the self-employed, a number of ECA countries offer preferential social tax treatment as an incentive for voluntary compliance and compensation for higher compliance costs. As Leibfritz (2011) points out, such an approach creates a distortionary element in the tax regime and encourages employees to change their status from dependent employment to self-employment.[116] Even without the explicit objective of reducing the social tax burden, tax policy makers face difficulties in applying the general social contribution regime to self-employed operating in a presumptive tax regime. While the level of social tax payment is generally a fraction of the net income of the self-employed, a presumptive tax regime does not produce any information on the net business income, and the requirement to calculate net income only for social tax purposes would conflict with the simplification objective of the

[115] The EU definition of a micro business refers to businesses with fewer than 10 employees and an annual turnover and/or annual balance sheet of not more than €2 million.

[116] For a detailed discussion, see Leibfritz (2011) and Packard, Koettl, and Montenegro (2012).

presumptive regime. An alternative calculation method in this situation is to apply the minimum wage as a tax base for the social tax, irrespective of the actual income of the self-employed. This was the case in Hungary under the EVA system and is applied for turnover tax payers in Poland. The result of this approach is that the self-employed tend to have a much lower tax wedge than regular employees, while incentives for salaried employees to become self-employed contractors (at least on paper) increases (see, for Poland, OECD, 2008a). Such a trend can be observed in Ukraine, where it is assumed that a remarkable increase in the number of unincorporated small businesses has been caused largely by salaried employees who register as small entrepreneurs and pretend to operate as independent contractors in order to secure the benefits of the simplified tax system (STS).

Table 6. Labour tax compliance times in ECA countries (hours)

Albania	94	Hungary	146	Romania	102
Armenia	162	Kazakhstan	70	Russia	76
Azerbaijan	97	Kosovo	41	Serbia	126
Belarus	88	Kyrgyz Republic	71	Slovak Republic	62
Bosnia	81	Latvia	139	Slovenia	96
Bulgaria	256	Lithuania	85	Tajikistan	48
Croatia	96	Macedonia	56	Ukraine	140
Czech Republic	217	Moldova	94	Uzbekistan	69
Estonia	34	Montenegro	98		
Georgia	56	Poland	124		

Source: PwC (2014)

Comparing small business development in Ukraine and the Russian Federation, an OECD analysis finds that the average unincorporated small business in Ukraine employs only approximately 1.6 persons. This figure has been falling, slowly but steadily, over the last six years. In Russia, by contrast, the average urban unincorporated small business in 2004 employed approximately 4.5 persons and the average number of employees had risen. The fact that so many Ukrainians who appear to work entirely alone have registered as individual entrepreneurs reinforces the perception that many are not actually self-employed entrepreneurs at all, but are seeking to exploit the benefits of the STS. A similar phenomenon may be observed in the small companies sector: the average number of employees in small companies fell from around eight to 6.4 persons between 2000 and 2006 (OECD, 2007).

A parallel but somewhat different issue is the social tax treatment of salaried employees in small businesses. Given the fact that social benefits such as health insurance, unemployment insurance and pensions are generally linked to the duration and level of contributions to insurance and pension schemes, the only transparent and reliable compliance method is for small business employers to calculate and transfer the precise employer contributions to these schemes and deduct employee contributions from salaries paid. There is thus no difference in the approach between small and larger business employers. For obvious reasons, this approach is not favoured by small

business operators using presumptive tax regimes to comply with their business income tax obligations.

Turnover tax regimes are biased against taking on formal sector employees, because – unlike the standard tax regime – the costs of hiring labour cannot be deducted as expenses, which effectively increases the small business tax burden. This bias is heightened by a social tax compliance burden which may equal, or even exceed, the presumptive tax compliance burden. While simple turnover-based presumptive tax regimes are expected to promote business formalization and migration out of the shadow economy for the actual business entity, they simultaneously create incentives for labour to move into the shadow economy.

Some ECA countries have tried to find ways to mitigate this risk and reduce the social tax compliance burden. A typical approach is to integrate social tax into the presumptive tax regime. In Latvia, the micro business turnover tax of nine per cent exempts the business from withholding employee personal income tax (general personal income tax rate of 25 per cent) and includes both employer and employee mandatory contributions to the social security system (employer share of 24.09 per cent; employee share of 11 per cent of income). The system works in its basic design for businesses of up to five employees. In the case of larger staff numbers, an additional two per cent on turnover is charged for each additional employee. Also, if the monthly salary of an employee exceeds €700 ($792), the excess amount is taxed at a rate of 20 per cent. In Ukraine, a business paying the unified tax of six per cent on turnover does not have to comply with income tax, social security, property tax, and some local tax payment obligations. Unified tax revenues are allocated based on a fixed ratio: 43 per cent of revenues is considered as the tax share, while 42 per cent goes to the state pension fund, and 15 per cent to the state social security fund.

The major difficulty with this approach is the lack of a clear attribution of the social tax payment to the beneficiaries of social benefits. In particular, when transfers to social agencies do not depend on the number of employees hired, a relationship between benefits and payments cannot be established. It may also be difficult for individual employees to prove that they have acquired social benefits while working for employers that are presumptive taxpayers. There is thus no real alternative to imposing regular compliance with the social contributions system on small businesses in the presumptive tax regime. A certain incentive for remaining in the formal labour system can be provided, as in the case of the Russian simplified tax system, by allowing the deduction of payments made to social security agencies from taxable turnover.

Approaches to taxing micro businesses

Defining the segment

Recognizing the fact that both the growth potential and the compliance capacity of micro businesses is substantially lower than for small businesses, many ECA countries have introduced specific tax instruments for the micro business segment. Thresholds defining the micro segment are frequently in the range of US$ 15,000 to 80,000, with Russia being the major outlier.

Table 7. Micro taxpayer thresholds and tax treatment in selected economies

Distinguishing micro and small businesses		
Country	**Micro business turnover threshold**	**Micro business regime (income tax treatment)**
Albania	ALL 2 million ($19,100)	Patent
Armenia	AMD 6 million ($14,750)	Patent
Bulgaria	BGN 50,000 ($34,700)	Patent
Croatia	HRK 149,500 ($28,900)	Patent
Hungary	HUF 6 million ($26,000)	Patent
Georgia	GEL 30,000($18,000)	Exemption
Kazakhstan	KZT 3,732,000 ($23,700)	Patent
Kosovo	€5,000 ($6,700)	Patent
Kyrgyz Republic	KGS 4,000,000 ($78,200)	Patent
Latvia	€50,000 ($56,500)	Patent
Macedonia	MKD 3 million ($66,000)	Exemption
Poland	Depends on number of employees	Patent
Russian Federation	RUB 60 million ($1.7 million) + not more than 15 employees	Patent
Serbia	SRD 6 million ($71,700)	Patent
Tajikistan	TJS 100,000 ($20,800)	Patent
Ukraine	UAH 1 Million ($110,000) + not more than 15 employees	Fixed single tax

Source: Authors, based on World Bank Group country reports.

Table 7 shows that not all systems have managed to limit the application of micro business regimes to very small entities operating around subsistence levels. Analysis of experiences in the region shows that one risk of micro business segmentation is that the very simple (and often very preferential) micro regimes also become an attractive model for businesses above the size of a micro business, and pressure to extend the regime up to the VAT threshold level is building.

The Serbian experience provides a good illustration of this dynamic. The patent regime in Serbia was initially targeted at micro operations at a level below SRD 2 million turnover ($23,900). However, the threshold was soon increased to SRD 3 million and, in 2013, the system was extended further to businesses with a turnover of up to SRD 6 million ($71,700). In principle, the system is targeted at "Any sole proprietor who, in

view of circumstances, is unable to keep books" (Article 40 of the Income Tax Law); however, it is also accessible to the well-educated and self-employed. The 2013 system reform at least managed to deny patent regime access to accountants, auditors, tax advisors and marketing agencies. However, an initiative by the Ministry of Finance also to exclude doctors and lawyers was not well received by lawmakers and was rejected by parliament. On the other hand, some countries, have demonstrated that, despite such pressure, a better alignment of the system threshold with the concept of targeting micro businesses may be feasible. An example is Tajikistan, which has limited the application of its micro regime from a previous turnover of $41,600 to $20,800 in 2014.

The micro business tax regime

The standard approach to micro business taxation in the region is the application of a patent regime. Frequently, these regimes are administered by local governments, and revenues go to local budgets. Only in a few cases, such as in Georgia, are micro businesses exempt from income taxation. Patent fees are generally determined by local governments, while national tax laws provide the basic structure for the regime and determine minimum and maximum patent rates. The actual design of patent regimes varies considerably in practice. In a few countries, a very basic patent regime has been put in place, with a fixed amount for all micro businesses, irrespective of business type and location.

Table 8. Patent regimes in Albania, Hungary and Kosovo

Country	Patent amount
Albania	ALL 25,000 annually ($240)
Hungary	HUF 50,000 monthly ($2,600 annually)
Kosovo	€37.50 quarterly ($200 annually)

Source: Authors, based on WBG country reports.

Much more frequent is the development of a detailed list of micro business activities and the determination of individual patent fees for each activity. A typical example of such an approach is the micro business regime in Bulgaria.

Bulgaria has experimented with its patent system for quite some time, at both national and local levels. While patents were initially administered by the national tax administration, the patent regime was transferred to local governments in 2008. The Law on Local Taxes and Fees lists 40 different categories of activity and specifies a minimum and maximum amount per established indicator. Local governments then determine the applicable amount for businesses in their territory, taking into consideration the precise location of the business, its economic importance, population

density in the area in which the business has been established, and the seasonal or permanent nature of the activity carried out.

Table 9. Patent regime in Bulgaria

Patent rates for artisans in Bulgaria (annual fee in Lev) *(Exchange rate: BGN 100 = $66.2)*			
Activity	**Indicator**	**Min.**	**Max.**
Retail trade (up to 100 m^2)	Square metres of sales area	2	20
Shoemaker	Tax determined considering the location of business premises	40	120
Repair of computers	Tax determined considering the location of business premises	300	1,300
Fortune teller	Tax determined considering the location of business premises	2,000	5,600
Hotels (1 & 2 stars) up to 20 rooms	Per room in consideration of location of establishment	25	250
Bowling hall	Per bowling alley	40	140
Restaurant (1 & 2 stars)	Per seat in consideration of location of establishment	1	35
Restaurant (3 stars)	Per seat in consideration of location of establishment	6	60

Source: Law on Local Taxes and Fees of the Republic of Bulgaria

Similar approaches are followed in a number of other Eastern European and FSU countries. The downside of such an approach is that the patent list becomes extremely voluminous and complicated. Patent lists distinguishing more than 100 categories of micro business activity are not uncommon. The Polish *Tax Card* system and the Latvian patent regime are examples of such approaches. In the Kyrgyz Republic, the patent list for the city of Bishkek includes 125 main categories of patent activity with an extensive list of sub-categories distinguishing, for example, whether a business stall is located in the first, second, or third lane of a local market. While such an approach intends to better align the patent amount with the presumed income of the micro business, the design of the regime risks creating a number of implementation problems:

- *Distinguishing activities of similar nature:* Many detailed patent lists could be simplified. In particular, the same patent rates are set for a variety of activities of similar nature, and thus unnecessarily complicate the list. The patent list for Bishkek, for example, includes eight different activities in the area of clothes manufacturing (activity list nos. 77-84); however, the patent rate for all eight activities is exactly the same. Merging similar activities into a single category is an easy option to simplify convoluted systems.
- *Mixed activities:* The more detailed the activity list, the higher the probability that a micro business carries out more than one listed activity. The Bishkek example

given above includes separate patents for the manufacturing of working clothes, other clothing and accessories, and underwear. In cases like this, micro businesses that are not highly specialized face the risk of multiplying their tax liability without necessarily achieving a higher profit than businesses that are active in only one area. There is therefore no clear justification for creating such a risk.

- *The gap issue:* Even a very extensive activity list can never be comprehensive. In fact, the more detail-oriented the list, the higher the probability that some micro business activities are missing. For affected business operators, it is then unclear how their patent rate will be determined or whether they can access the regime at all.
- *Determination and update of patent rates:* It is a major exercise to set more than 100 different rates and ensure that they are properly updated and reflect changes in the business environment, profitability and costs. Ministries of finance and tax administrations are typically not well equipped, nor do they have the required resources to perform this task properly. Sufficient data and information to allow a thorough calculation of appropriate rates for very specific activities is rarely available.
- *Acceptance and fairness:* As a result of the rate-setting problems, international experience shows that the acceptance level of detailed patent regimes in the micro business community is low. Business operators tend to consider the rate-setting process to be non-transparent. They also perceive the patent rate that they have to pay as arbitrary and unfair.

Table 10. Service businesses in patent category 111 (computer and copier material processing) operating in Bishkek City

Annual turnover	**KGS 100,000**	**KGS 500,000**	**KGS 1 million**	**KGS 3 million**
Tax burden under patent regime	12,000	12,000	12,000	12,000
Tax burden under simplified tax regime	6,000	30,000	60,000	180,000
Tax burden under standard regime (assuming a profit margin of 30%)	Income tax: 3,000 Sales tax of 3.5%: 3,500 TOTAL: 6,500	Income tax: 15,000 Sales tax: 17,500 TOTAL: 32,500	Income tax: 30,000 Sales tax: 35,000 TOTAL: 65,000	Income tax: 90,000 Sales tax: 105,000 TOTAL: 195,000

It is therefore desirable to streamline the detailed patent lists and categorize micro business activities more broadly. A reform in this direction has been implemented in Tajikistan.

Improving the operation of a patent regime: micro business tax reform in Tajikistan

Tajikistan operates a patent system for individual entrepreneurs with a turnover of up to SM (Tajik somoni) 100,000 ($20,800) annual turnover and size of business premises up to 30 square metres. The cost of the patent is fixed according to type of activity, and varies from SM 240 ($50) per month for transportation of oil, liquid gas, and cement by specialized transport, to SM 30 ($6) for transportation by motor-scooter. The patent was initially only valid for one narrowly specified category of activity and had to be renewed annually. A reform of the patent system in 2008 introduced major simplifications. First, the new patent system offers an expanded scope of permitted economic activities. It reduced the number of patent categories from 49 patents (with 70 sub-activities) to 28 (covering 169 sub-activities). Second, tax inspection of patent holders is limited to verification that the individual entrepreneur has a patent, that the business activity is allowed by the patent, and that the patent holder has a receipt of payment for the patent. Previously, the patent holder was required to keep a cash register and maintain transaction logs, on the basis of which the tax inspector would determine tax obligations. Third, the patent no longer has to be updated annually. The patent system has become much more attractive since this reform, and 15 per cent more patents were issued in 2008 than in 2007 (IFC, 2009a).

Use of additional indicators

While a general patent regime establishes the same tax liability for all micro businesses active in a specific field, such an approach might be considered inappropriate, as the prospect of earning income depends critically on certain criteria, such as the business location or facilities. Some patent regimes have therefore introduced a small number of critical indicators to take account of the specific business situation. By far the most popular and, at the same time, most transparent indicators are the number of employees and the location of business premises. One example of a system combining the use of these indicators is the *Tax Card* regime in Poland.

The use of other indicators tends to be much more problematic. This is the case particularly where turnover elements have been introduced to create a progressive patent regime. In Croatia, for example, the lump sum tax amount depends on the level of the micro business turnover. The micro business in this case is still required to maintain a cash book and calculate its annual turnover. With regard to calculation of the tax liability, a normal turnover tax could obviously be applied in this case, which would eliminate the jump in the tax burden arising from moving from a higher turnover band.

Table 11. Poland: Monthly patent rates for selected micro-business activities for 2014 (in US$)

Type of business	No. of employees	Business location (number of inhabitants)		
		Less than 5,000	5,000–50,000	More than 50,000
Jewelry shop	0	136	149	149
	1	232	266	266
	2	348	388	388
Barber shop (ladies)	0	38	45	51
	1	103	116	129
	2	143	161	172
	3	172	185	201
	4	185	201	218
	5	225	257	291
Flower shop	0	161–218	177–278	218–356
	1	177–281	218–356	278–406
	2	218–356	278–406	375–507

Table 12. Calculating the micro business tax liability in Croatia (in HRK)

Business turnover	Tax liability
Up to 85,000 ($14,900)	12,750 ($2,230)
85,001–115,000 ($20,100)	17,250 ($3,000)
115,001–149,500 ($26,100)	22,425 ($3,920)

Patent system stability and predictability

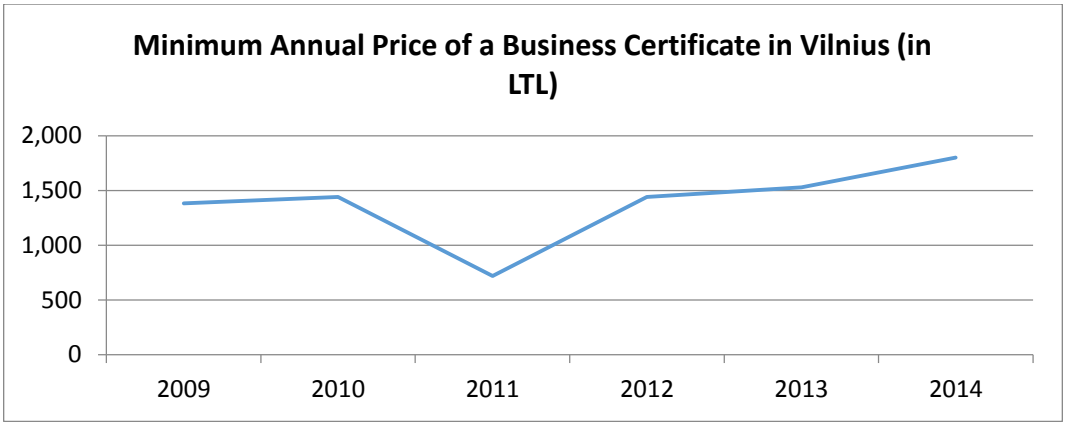
While patent systems have the potential to provide a transparent approach to micro business taxation and reduce micro business compliance costs, the unpredictability of tax rates can be a major issue in practice, particularly when patent rates are set by and changed based on the revenue needs of local governments. In Lithuania, for example, a country where the patent regime extends to small businesses up to the VAT registration threshold, patent rates are determined annually by local councils, and the patent levels change substantially. Although an important revenue source for local governments, the unpredictability of the regime may add further disincentives to business formalization.

Abuse risks of micro regimes

For small businesses in the higher turnover range, patent-based taxation may be an easy and attractive way to reduce the overall tax liability. In principle, patent regimes may therefore create a risk of downward migration. In practice, such a risk can be largely mitigated by (i) an appropriate limitation of the application of the micro regime, and

(ii) a design of the small business regime taking into account small business profitability and ability to pay.

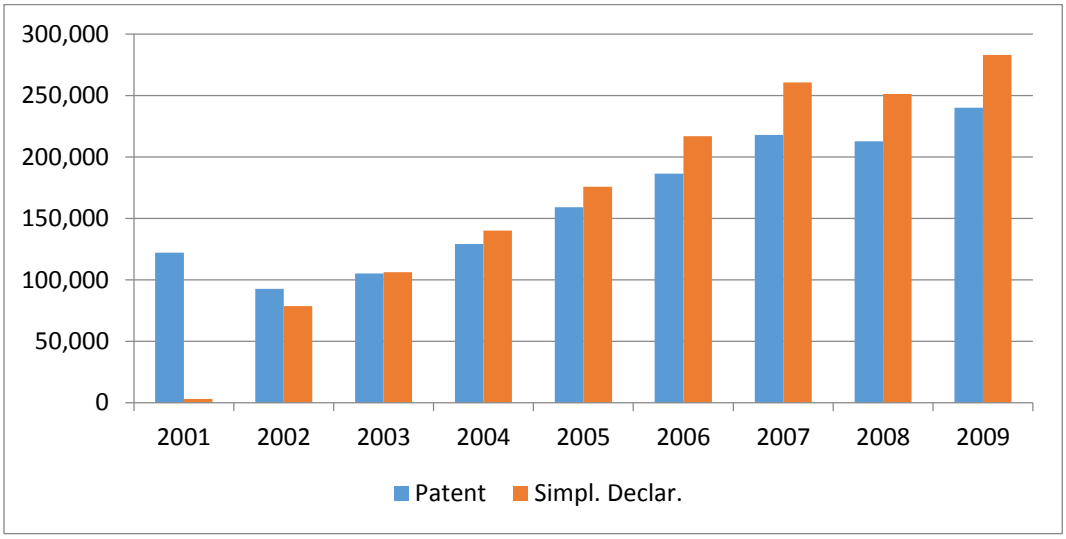
Figure 5. Volatility of patent rates in Lithuania



Source: Vilnius City Council data.

In the case of Georgia, which exempted micro businesses from taxation, initial assessments of taxpayer behaviour following the reform suggest there was no noticeable downward migration of small businesses (Bruhn & Loeprick, 2014). Downward migration is thus not a given, provided that the micro threshold is sufficiently low while effective compliance management covers the segment and the small business regime is operated reasonably well.

Figure 6. Number of taxpayers applying simplified tax regimes in Kazakhstan



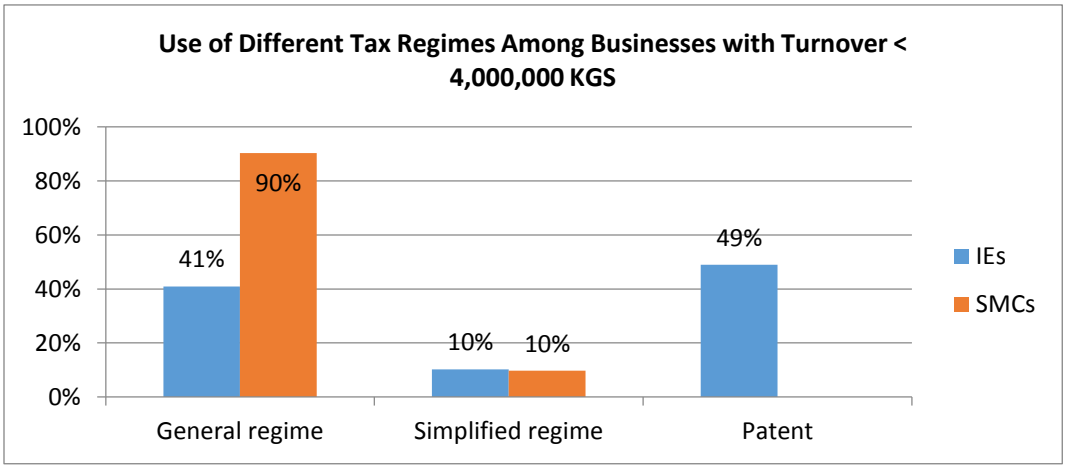
Source: Kazakh Revenue Service

The risk of system abuse increases with high patent-system thresholds. In Kazakhstan, the upper threshold of the micro business (patent) regime of a turnover of around KZT 4 million[117] ($25,400) extends to net incomes above the average annual wage of KZT 1.6 million[118] for businesses with profit margins above 40 per cent. The regime is very popular, with ample anecdotal evidence that high-income self-employed use the regime to lower their tax liability.

A patent regime may thus erode the small business tax regime and create major disincentives for small business growth when it reaches out to businesses up to the VAT registration threshold. This can be observed in Bulgaria, the Kyrgyz Republic and Lithuania.

In the Kyrgyz Republic, small businesses have the choice to apply either the turnover-based small business tax regime (simplified regime) or the micro business patent regime, with the patent regime generally offering a comparatively favourable tax treatment, as demonstrated in Table 13. As a result, acceptance of the simplified regime is minimal and most small businesses request patents. As the patent regime does not require any books and records, business growth and migration into the VAT system and the standard income tax regime becomes less likely, while the tax administration faces difficulties in controlling abuse of the regime effectively.

Figure 7. Use of tax regimes in Kyrgyz Republic



Source: IFC SME Survey, 2009

[117] Two hundred times the minimum monthly wage of KZT 19,966.
[118] The average monthly nominal wage in December 2013 was KZT 137,043.

Small business taxation

Defining a small business system threshold

International good practice recommends aligning the upper threshold of a simplified small business tax regime with the VAT registration threshold. Businesses registered for VAT must comply with advanced accounting standards, and should therefore also be in a position to calculate their net income easily for income tax purposes. In addition, VAT compliance management benefits from the possibility of cross-checking information from income tax and VAT returns. However, using the VAT threshold as a criterion to determine the ceiling of the simplified small business tax regime is less evident in cases where the VAT threshold is either very high – thereby limiting the application of VAT to large businesses – or very low. In the first scenario, the simplified regime would extend to the medium business segment, and thus to a group of businesses which are able – and should be obliged – to calculate their net income; while, in the second scenario, a low VAT threshold also impedes compliance cost reductions on the direct taxation side.

One notable feature of presumptive tax design in the ECA region is a frequent lack of proper system coordination between the presumptive tax regime and the VAT regime, in particular with presumptive regimes extending to medium-sized businesses required to register for VAT, as in Belarus, Macedonia and Ukraine.

Approaches to turnover-based small business taxation

Turnover has become the most widely-used base for small business tax systems, and its design varies considerably from country to country. The two key system design alternatives are:

- Presumptive income tax regimes replacing only personal income tax or corporate income tax, versus single tax regimes offering small businesses the option to pay only one tax rather than a variety of taxes.[119]
- Single tax regimes with a single tax rate for all types of small business (as in Azerbaijan, with regional tax rate variations), versus multiple rate regimes using different tax rates for different business categories to account for different (assumed) average profit margins, particularly in the trade versus service segment.

[119] If a general VAT exemption threshold is in place, typically no VAT component will have to be included in the presumptive tax amount. Small businesses will not be able to recover VAT on inputs purchased, and should be required to comply with regular income tax obligations if they decide to register voluntarily for VAT. If a presumptive VAT payment is determined for taxpayers below the threshold, the associated filing and payment obligations should be aligned with the presumptive income tax requirements.

Table 13. Design characteristics of turnover tax regimes in the region

Country	Single rate regime	Multiple rate regime	Turnover rate	Replaces income / profit tax only	Single tax regime	Application to incorporated businesses
Armenia		x	3.5% for production and trade 10% for rental income 12% for notaries 5% for other activities		Replaces income tax and VAT	Yes
Azerbaijan	x		4% for businesses in Baku 2% for other businesses		Replaces income tax, property tax and VAT	Yes
Belarus	x		5% in case system also replaces VAT 3% in case VAT is paid separately		Unified tax replacing general tax regime	Yes
Georgia	x		5% (possibility to reduce to 3%)	x		No
Kazakhstan	x		3%		Unified tax including social security	Yes
Kosovo		x	3% for trade and transport 5% for services, professional activities, entertainment	x		Yes
Kyrgyz Rep.		x	4% for trade 6% for other activities		Profit tax and sales tax	Yes
Latvia	x		9% tax rate increase if more than five employees by two additional percent points per employee		Income tax, social insurance payments, state duty on entrepreneurship risks for employees	Yes
Macedonia	x		1%	x		Yes
Romania	x		3%	x		Yes
Russia	x		6%		Single tax regime, also replacing social taxes	Yes
Tajikistan	x					
Turkmenistan	x		2%			Yes
Ukraine	x		3% when also includes VAT 5% without VAT		Income tax, land tax, trade patent fee, social security and pension contributions, some local taxes	Yes
Uzbekistan	x		6% with different rates for specific areas, such as, for example, 5% for wholesale trade		All government taxes and local taxes and charges, except local fee for commercial activities	Yes

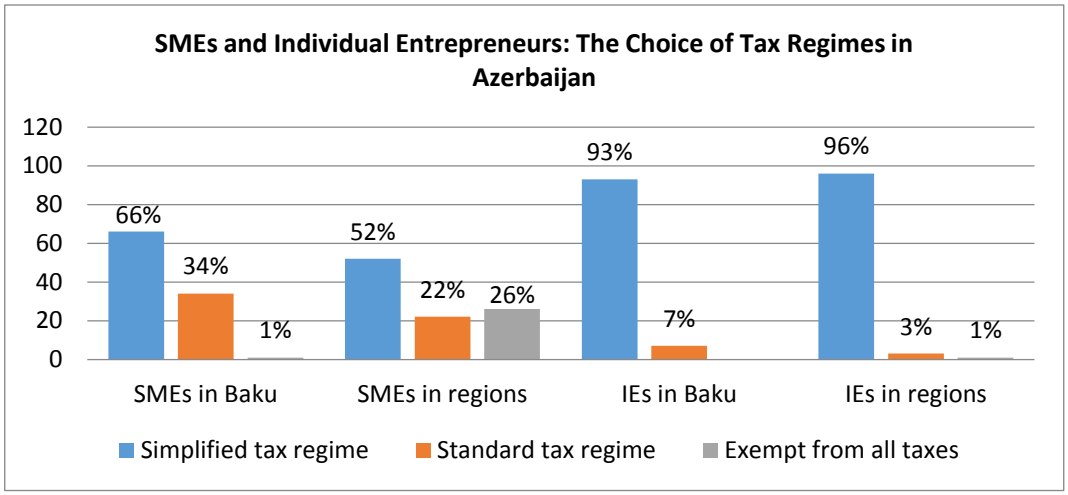
Use of presumptive regimes: lessons learned

Low take-up and the design of appropriate rate structures

Two conclusions can be drawn regarding the use of presumptive tax regimes. First, experience shows that newly-introduced presumptive regimes are not automatically widely applied by small businesses. Regime take-up rates tend to be low initially, and only increase as a result of taxpayer information campaigns and trust being built in the small business community with regard to the practical application and potential benefits of the regime.

Second, the use and popularity of presumptive regimes is not directly correlated with the turnover rate applied. Turnover tax regimes with a low tax rate do not automatically attract more businesses than regimes with a higher tax rate. Azerbaijan is an example of a country applying different turnover rates depending on the location of the business: while a four per cent rate on turnover is applied for businesses located in Baku, businesses in the regions only have to pay a two per cent turnover rate. Survey results suggest that this rate differentiation has no measurable impact on the level of system use.

Figure 8. Use of tax regimes in Azerbaijan

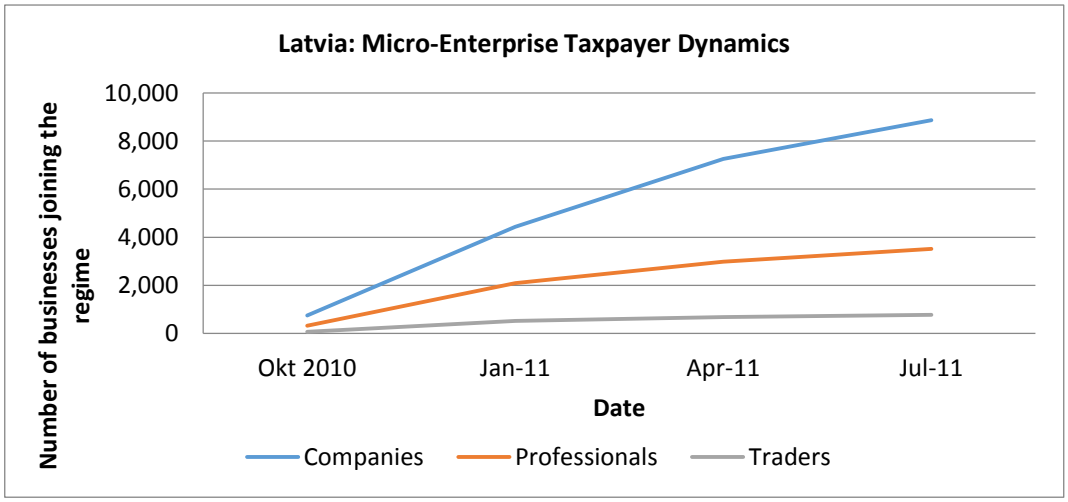


Source: IFC (2009b).

The level of the tax burden as a percentage of turnover may impact on the composition of the presumptive taxpayer group. A system with higher turnover tax rates may become less attractive for businesses with low profit margins. This is an unwelcome result, as businesses with lower profit margins, which frequently operate in the retail trade, generally have less capacity to comply with net income taxation than service businesses and professionals. The system thus risks being biased toward segments of the small business population that have less need for compliance simplification. An example of such a phenomenon is the micro-enterprise tax regime in Latvia, which has a comparatively high rate of nine per cent on turnover. In practice, this has translated into a situation in which the regime has

generated very little interest among small traders. Its primary users are self-employed professionals and, in particular, small limited-liability companies.

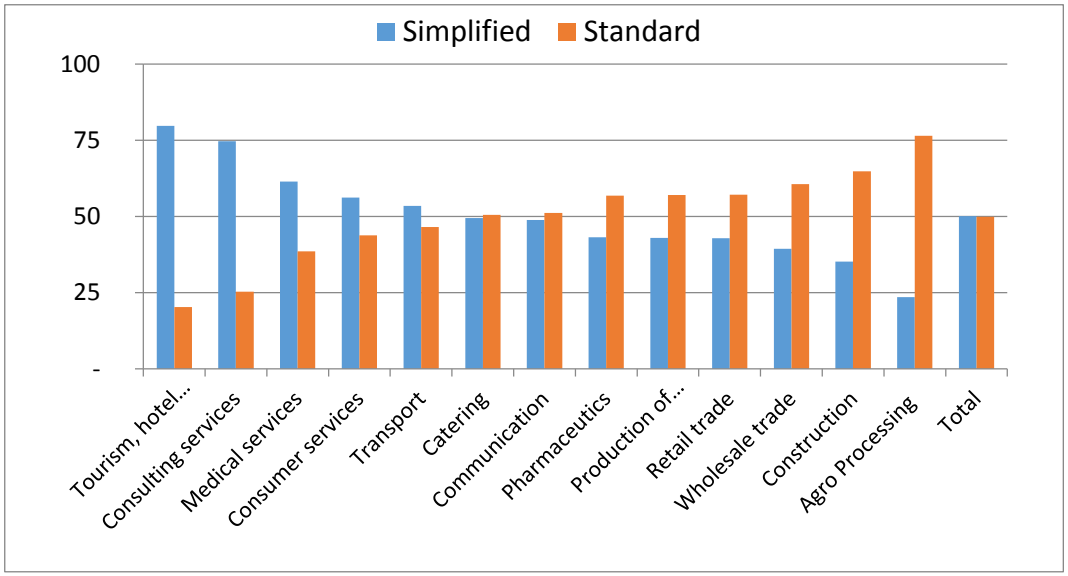
Figure 9. Latvia: Businesses opting for presumptive taxation by type of business



Source: Bruna & Sneidere (2011), with additional data from the Revenue Service.

Interestingly, the same phenomenon can be observed even in countries with a much lower turnover tax rate, as shown in the example of Tajikistan, which operates only a four per cent turnover tax (five per cent for turnover above SM 200,000). While the turnover tax regime is applied by 80 per cent of small tourism operators, 75 per cent of small consultancy service providers, 61 per cent of medical service providers, and 67 per cent of consumer service businesses, the take-up rate is only 39 per cent for retail businesses and 35 per cent for construction companies.

Figure 10. Tajik Firms in the simplified versus standard regime by sector



Source: IFC SME Survey, 2008

A gradual change in the composition of businesses using the simplified tax regime has also been experienced in Hungary with the operation of simplified entrepreneurial tax (EVA) and the gradual increase of the EVA tax rate. EVA was introduced in Hungary in 2003 as a flat tax paid on sales revenues. Eligibility was limited to private entrepreneurs and business entities that had been in business for at least two years, had an annual income (including VAT) not exceeding Ft 25 million (approximately $115,000), and with individuals as owners. EVA payers were not subject to income tax, dividend tax, company car tax or VAT. The EVA rate was initially one per cent, levied on turnover with some minor adjustments. EVA quickly became very popular, and the number of businesses applying for taxation under the simplified tax increased from 59,000 in 2003 to 83,000 in 2004 and close to 100,000 in 2006, when the tax rate was increased to 25 per cent. Surprisingly, the participation level remained stable and still covered around 96,000 businesses in 2010. As the system was attractive primarily to businesses with margins exceeding 60 per cent, it seemed to be used extensively by engineers, lawyers and accountants, who operate with low cost/income ratios, implying that EVA reduced their tax burden. Some entrepreneurs used EVA even if it entailed a higher tax burden because it reduced their administrative costs (OECD, 2008b, based on Semjén, Tóth, & Razakan, 2008). The EVA rate was further increased to 30 per cent in 2010 and to 37 per cent in 2012, before being abolished in 2013.

The risk of factual discrimination of low-profit trading businesses can be avoided by introducing a rate differentiation into the turnover tax regime. Small businesses in the trade segment can be offered a lower turnover tax rate, while service businesses and, in particular, self-employed can be taxed at higher rates. While such an approach is still a rarity in the region, it has been tried successfully by some ECA countries. Poland, which has developed the most sophisticated rate differentiation, demonstrates that the benefits of this approach are twofold: it eliminates system entry barriers for businesses in retail and wholesale trades, and improves presumptive tax revenue generation, as higher turnover rates on high-profit segments considerably increase overall presumptive tax collection.

Table 14. Registered small businesses in the Polish turnover tax system

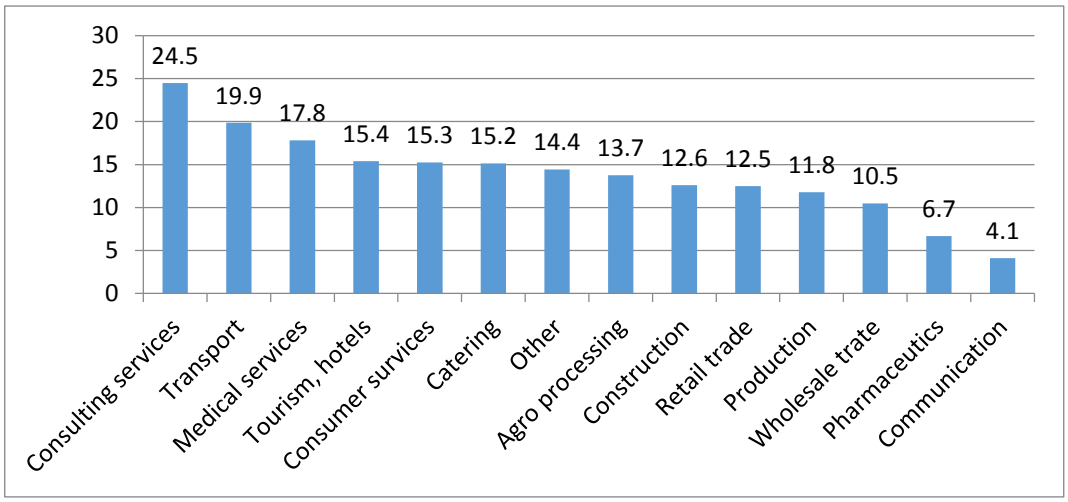
Category	Turnover rate (%)	Number of taxpayers	Tax revenues (million $)	Tax collection per business ($)
Trade and catering	3	177,837	77	433
Manufacturing, construction, transport	5.5	178,302	274	1,537
Service businesses	8.5	370,715	257	693
Car rental services	17	4,554	5	1,098
Liberal professions	20	56,292	72	1,279

Source: Grabowski (2011).

The major drawback of applying the approach in practice is the lack of reliable data on average small business profit margins to justify the scope and level of rate differentiations. Data from statistical offices in ECA countries have frequently

proved insufficient as a basis for rate differentiation, and attempts to collect information on average profit margins in various small business segments through targeted taxpayer surveys have generally not provided sufficiently reliable data. Consequently, such survey results are difficult to interpret and provide only a limited basis for policy decisions. An example of such a situation is the IFC SME survey conducted in Tajikistan in 2009. This survey shows major differences in profit margins within the service segment, while most activities listed generally only show a relatively small overall profit margin ranging between 10 and 15 per cent. This does not explain the considerable differences in simplified system use discussed above, and probably indicates that real profit margins are substantially different from the margins provided for the survey.

Figure 11. Small business profit margin analysis in Tajikistan



Source: IFC SME survey. 2009

A reliable profit margin differentiation for tax purposes requires input from tax administrations, particularly through the collection and analysis of SME tax audit data, and ideally as a combination of targeted risk-based audits, as well as a number of random audits conducted to verify margin estimates. However, such audit data are not generally available, or at least have not yet been systematically analyzed.

With an incomplete and unreliable basis for turnover rate differentiation, an approach such as in Poland, which distinguishes five small business segments, becomes questionable. More appropriate in this situation would be a simpler two-rate system, as practiced in Kosovo and the Kyrgyz Republic. A further alternative to consider might be the approach introduced in Georgia in 2011. Here, the small business tax regime for businesses with a turnover of up to GEL 100,000 ($57,500) has a standard rate of five per cent for small businesses in all segments; however, a business can benefit from a reduced rate of three per cent if it can document, by presenting corresponding invoices, that business expenses amount to at least 60 per cent of turnover. This rate differentiation is not directly linked to the business activity, but is supposed to reflect the actual business profit margin. Undoubted

disadvantages of this approach are the need for the tax administration to verify invoices presented by small businesses and the risk of disputes about the tax recognition of these invoices, but it provides an innovative method for increasing the fairness of a turnover-based presumptive tax regime. A similar approach is currently being discussed for the trade segment in Armenia.

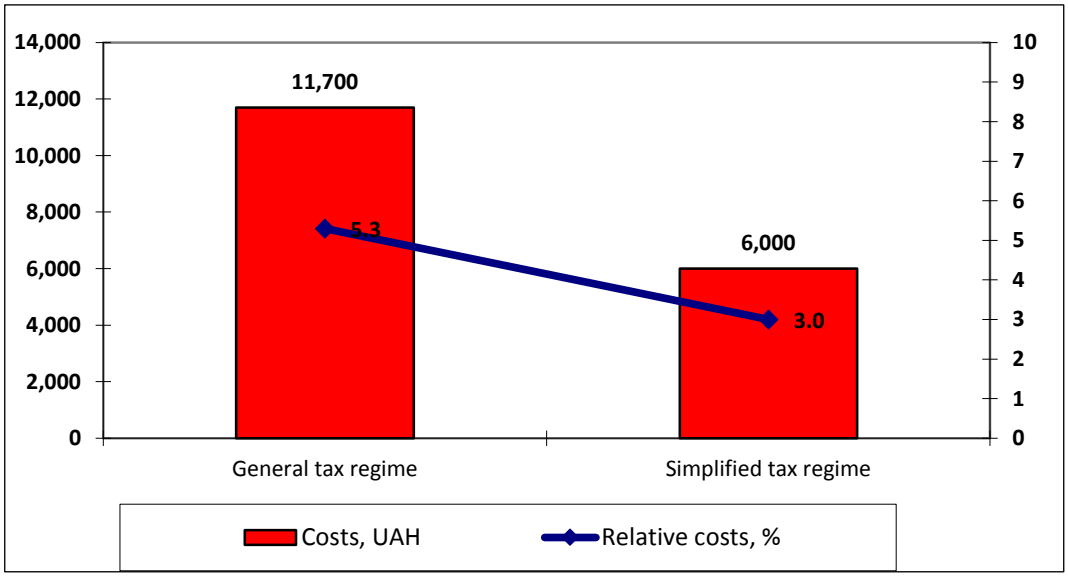
Presumptive regimes, the business environment and compliance management

Surprisingly, despite the extensive use of presumptive tax regimes in the region, analysis of the impact of the system is still sporadic. Of particular importance is impact analysis of tax compliance costs and business formalization.

Compliance cost reduction

Business surveys provide some indication of compliance costs associated with different tax regimes. A simple comparison of compliance costs in Ukraine, for instance, indicates a cost reduction of almost 50 per cent for small businesses operating in the presumptive regime.

Figure 12. Comparison of tax compliance costs in Ukraine in the presumptive versus standard tax regime

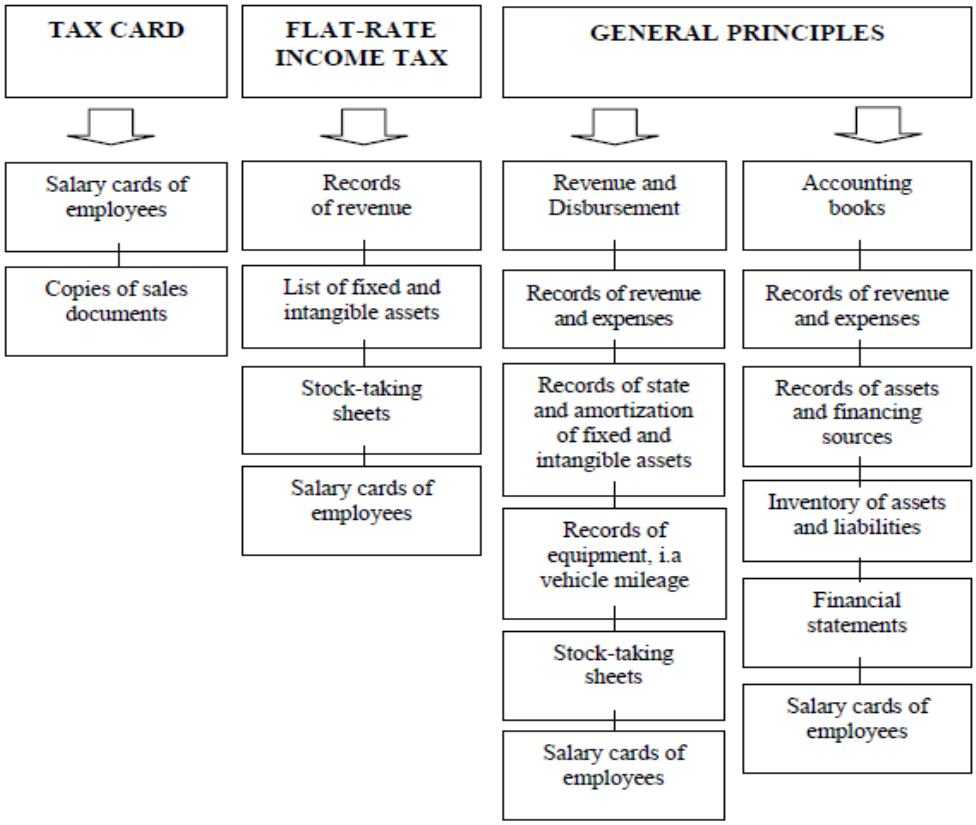


Source: IFC, 2009c.

Similar survey-based reviews consistently identify accounting and tax filing obligations as a major element of compliance costs, thus supporting general claims of cost reduction following the introduction of simplified accounting and filing requirements for presumptive taxpayers. Fully exploiting the cost-benefit potential of presumptive taxation therefore requires appropriate simplification of bookkeeping and reporting requirements. A full alignment of bookkeeping standards with the requirements of a turnover-based tax system would, in principle,

require the business only to keep records of its turnover. This approach, however, would not provide sufficient data for a presumptive taxpayer risk analysis and identification of businesses abusing the regime; therefore, tax administrations generally require more extensive documentation of business transactions. Nevertheless, compared with accounting requirements in the general tax regime, a major simplification of the accounting burden can be achieved.

Figure 13. Comparison of accounting obligations in the micro business (tax card), small business (flat rate income tax) and general tax regimes in Poland



Source: Jaworski, 2011a.

Small businesses also consistently highlight problems with the number of taxes with which they are required to comply. Armenian businesses, for example, cited the number of taxes as their most widespread concern with regard to the operation of the tax system (see IFC, 2011). In Belarus, businesses complained that a normal SME in the general tax regime has to pay on average 12 taxes and duties (Kireeva & Rudy, 2007). Single small business taxes, which combine various national (and, in some countries, local) taxes into one tax payment, therefore create additional compliance cost reductions for small businesses. This can be a major benefit in practice. The single tax in Ukraine, for example, substitutes 12 other taxes and duties in addition to income/profit tax, VAT and land tax.

Business formalization and compliance

Little thorough analysis is available with regard to the impact of simplified small business taxation on business formalization and compliance behaviour in the ECA region.[120]

Figure 14. Ukrainian businesses registered for unified taxation

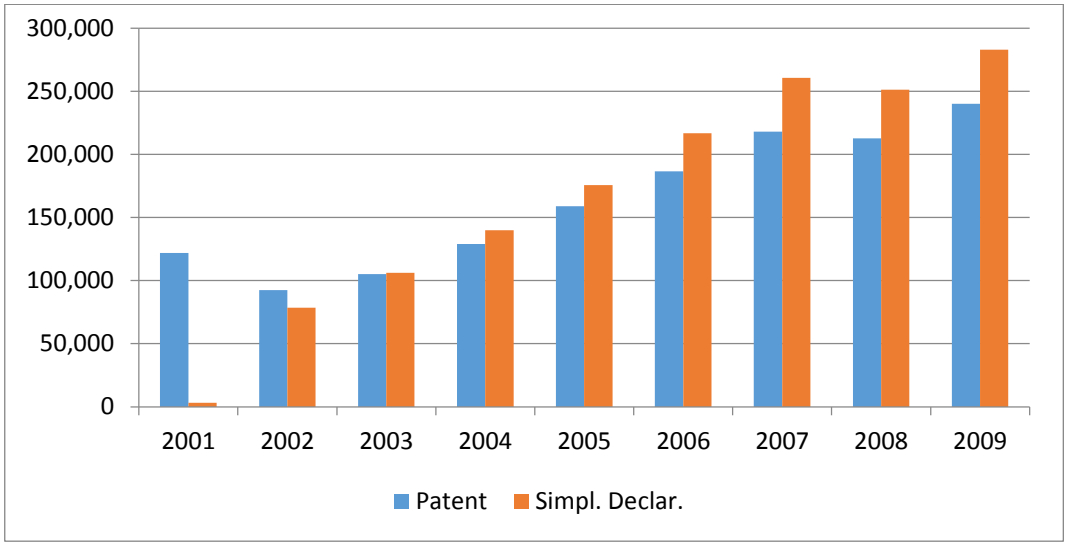
Source: World Bank (2006) and additional data from the Revenue Service.

At first sight, when considering registration trends, the experience in many ECA countries looks impressive. In the short term, presumptive systems may have a relatively low initial take-up rate, probably due to a lack of information on the part of small business operators about the regime and its benefits, or distrust in its stability and implementation. Generally, however, participation tends to grow remarkably over time.

A typical example of such a development is Kazakhstan where, since 2002, the number of taxpayers applying one of the simplified regimes has been growing annually by an average of 27 per cent in the case of the simplified declaration regime and by 18 per cent in the case of the patent regime. This growth rate exceeds the overall growth of the taxpayer population.

[120] Studies revealing a positive impact of presumptive taxation on business formalization have been conducted in some non-ECA countries, such as Brazil. See Fajnzylber, Maloney, and Montes-Rojas (2011), and Monteiro and Assunção (2006).

Figure 15. Number of taxpayers applying simplified tax regimes in Kazakhstan (2001-09)



Source: Tax Committee of the Ministry of Finance of the Republic of Kazakhstan.

However, such growth figures do not provide any information on the actual effects of presumptive tax regimes on the formalization decisions of businesses. What is generally lacking in ECA countries (and worldwide) is analysis of the previous tax situation of small businesses joining the regime. The presumptive taxpayer population includes (i) tax-registered small businesses changing their tax status and moving from the general to the presumptive regime (see also the example of Uzbekistan below); (ii) newly-established businesses registering with the tax authority and selecting the presumptive taxation approach; (iii) some – hopefully not too many – larger businesses migrating into the presumptive regime. It is therefore completely unclear whether and to what extent growth in registration can be attributed to informal businesses deciding to formalize.

Several analyses of the dynamics around presumptive tax regimes provide an indication that these have contributed to improved voluntary tax compliance. For Ukraine, Thiessen (2003) estimates that, looking at the years 1999 and 2000, the presumptive tax for small and medium-sized businesses has reduced the shadow economy by 11-14 per cent. More recently, the Russian SME Resource Center conducted a survey and analysis of businesses using the simplified taxation system in the Russian Federation (Mikhalkin & Alexeeva, 2009), concluding that adoption of the simplified tax system resulted in the legalization of about 30 per cent of SME income. Looking at the use and performance of simplified tax systems in the period 2003-07, the analysis finds that the number of legal entities and individual entrepreneurs taxed on a presumptive basis had almost tripled in 2007 compared with 2003 (the number grew from 692,000 to 1,602,000); at the same time, the revenue performance of the simplified tax regime increased by a factor of almost five, from Rub 17.2 billion in 2003 to Rub 84.1 billion in 2007. This means that the presumptive tax payment per small business more than doubled in the period

reviewed, from Rub 23,000 in 2003 to Rub 52,000 in 2007, indicating a greater willingness of presumptive taxpayers to legalize their business transactions and declare their actual turnover.

Similar revenue increases were reported in Ukraine following the introduction of a fixed tax regime, and the tax administration found that the amount of tax collected from businesses that decided to move into the fixed-tax regime quadrupled compared with the amount collected from the same group of businesses before the fixed tax was introduced (Semikolenova, 1999).

Presumptive regime threshold and business development distortions

While the existence of a simplified presumptive tax regime may have considerable benefits for the SME segment, multiple risks are associated with the impact of the presumptive regime on revenue generation and the integrity of the tax regime, as well as potential detriments to business growth and development. These detriments result, in particular, from a major tax burden and compliance cost increase for businesses migrating from the presumptive to the standard tax regime, and an undesirable incentive for larger businesses to migrate downwards into the presumptive regime.

As a result, country analysis tends to reveal the erosion of the general tax regime due to the attractiveness and popularity of presumptive tax regimes. This erosion needs to be balanced against the benefits of the regime, and may not always be as serious as might be assumed when glancing at anecdotal evidence. Sometimes, a skewed taxpayer distribution may simply reflect the predominance of the micro and small business segment in an economy.

Presumptive tax regimes for many small businesses offer the possibility not only of facilitating tax compliance and reducing compliance costs, but also reducing the actual business tax burden. There is no sound justification for such tax liability reduction, which violates the ability-to-pay principle;[121] in practice, however, many presumptive tax regimes result in a comparatively low tax burden compared with similar businesses operating in the standard tax regime.

Upward migration

Indeed, the percentage of small businesses taxed on a presumptive basis that eventually migrate voluntarily into the standard tax regime is low. However, this observation, as such, is insufficient to assume that the presumptive regime threshold establishes a business growth barrier. Such a conclusion would require the additional diagnosis that presumptive taxpayers tend to grow to a level close to this

[121] This situation is similar to the practice of offering lower corporate income tax rates to small companies, although in that case, at least an argument could be made that the lower tax rate compensates for higher compliance costs. This is not a valid argument in the case of presumptive taxation, as the presumptive regime is already supposed to align the compliance burden with the compliance capacity of the business.

threshold and then show no further increase in their turnover. Many small businesses tend to remain small and operate at the lower, not the upper, end of the presumptive regime turnover scale. In Ukraine, for example, looking at the distribution of unified taxpayers according to turnover levels, Alm and Saavedra (2006) find a remarkable upward migration within the system. Over five years, 2000 to 2004, the total number of individual simplified taxpayers more than tripled. However, while the increase in the lowest band has been modest (an increase of 69 per cent), taxpayers with payment obligations in the four highest bands (more than Hrv 200,000) increased by more than 12 times. This is only to a small extent due to inflation, as annual inflation rates were below 10 per cent in the years analyzed.

Table 15. Number of taxpayers and tax payments of physical persons in the Ukrainian simplified tax system

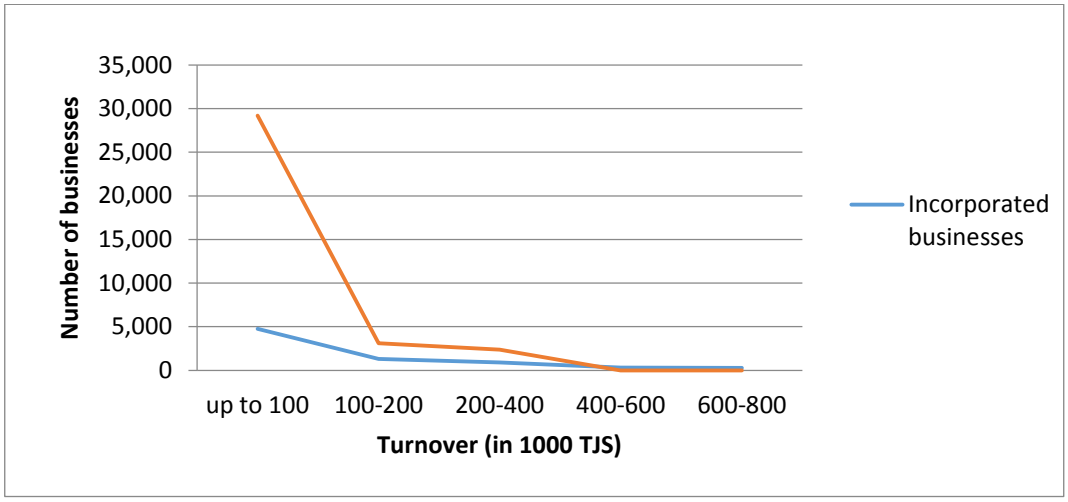
Annual tax payment (in UAH)	Number of taxpayers					
	2000	2001	2002	2003	2004	Increase (%)
0–2,400	88,077	146,351	173,765	162,227	148,541	68.6
2,400–7,200	26,797	39,950	56,421	72,355	73,369	173.8
7,200–55,000	44,512	73,356	113,767	163,902	188,158	322.7
55,000–100,000	8,681	15,472	26,292	41,043	52,297	502.4
100,000–200,000	7,593	14,065	25,035	40,458	53,046	598.6
200,000–300,000	2,918	6,110	11,853	20,294	28,277	848.6
300,000–400,000	1,664	3,524	7,736	13,974	20,456	1129.3
400,000–500,000	1,710	3,707	10,895	23,446	34,506	1917.9
Over 500,000	605	741	1,089	2,251	3,800	528.1
TOTAL	182,557	303,276	426,853	539,950	602,400	230

Source: Alm & Saavedra, 2006.

Still, more than two-thirds of presumptive taxpayers are grouped into the three lowest bands of the regime and do not operate even close to its upper threshold, suggesting that the large majority of small businesses in the regime are not concerned with a potential system threshold growth obstacle.

This review of taxpayer distributions does not imply that a system threshold barrier for small business growth does not exist. Rather, it indicates that the number of presumptive taxpayers affected by the system threshold is comparatively small. In addition, it is important to take into account that the presumptive regime threshold is not necessarily the only obstacle to business growth. Even without the need to abandon the preferential presumptive regime treatment, reaching the VAT threshold can become a major barrier to small business growth. A strategy to facilitate the transition from the presumptive to the standard tax regime therefore needs to include, as a core element, the facilitation of VAT compliance procedures for medium-sized businesses, in particular by offering VAT cash accounting schemes. A cash accounting option might also be considered for income taxation of medium-sized businesses.

Figure 16. Tajikistan: Number of taxpayers in simplified tax regime according to turnover bracket

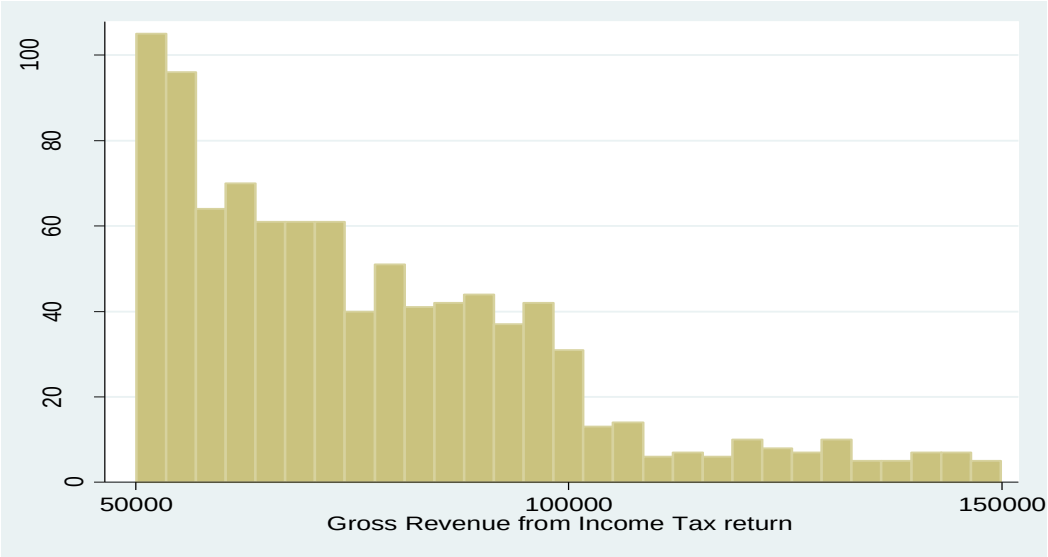


Source: Tajikistan Tax Administration, 2010.

While several Central and Eastern European countries have reduced the barrier for migration out of the presumptive regime by better aligning their small business tax regime with the general tax regime (see below), Russia, Ukraine and Belarus have considerably increased the eligibility thresholds of their presumptive regimes. In the Russian Federation, the threshold for application of the STS for incorporated businesses was raised in 2010 from RUB 26.8 million to Rub 60 million turnover, with the staffing threshold of a maximum of 100 employees remaining unchanged; similarly, in Ukraine, the presumptive regime threshold increased in 2012 for incorporated businesses from one million to five million UAH, with the maximum staffing of 50 employees also remaining unchanged.

Increasing the threshold largely eliminates the growth obstacle for businesses with a turnover below the old threshold, but does not provide an overall solution to the problem. As Alexeev and Conrad (2013) point out, a business at the RUB 60 million threshold may still face a RUB 400,000 tax increase as a result of a one RUB increase in turnover. This may also increase problems of system operation, and the risk of system abuse becomes more acute. Rather than raising the system threshold, an effective strategy for facilitating business migration should at least include the following components: (i) aligning the tax burden in the presumptive regime with the tax burden in the standard tax regime; (ii) imposing basic bookkeeping standards on presumptive taxpayers; (iii) introducing some compliance facilitation measures, in particular for VAT compliance, for medium-sized businesses; (iv) offering targeted taxpayer services for business migration. In addition, a risk-based audit approach aimed at identifying businesses that should be moved into the standard tax regime should be applied in order to avoid unfair competition for medium-sized businesses in the standard regime.

Figure 17. Distribution of firms in Georgia by turnover below and above the VAT threshold of 100,000 GEL[122]



Source: Bruhn & Loeprick, 2014.

Downward migration

In a number of cases, migration of businesses from the standard tax regime to the presumptive tax regime is perfectly legitimate. It may be a result of either a shrinking business turnover – making the business eligible for the presumptive regime – or a business that always qualified for presumptive taxation but opted to be taxed in the standard regime and subsequently changed its system selection.

Tax administrations frequently report incidences of widespread presumptive regime abuse by larger businesses. An extreme case is Ukraine, where the tax administration found cases of larger firms splitting into 20 or more small businesses, thereby qualifying them for the presumptive tax regime. As such, government officials are concerned that as much as 50 per cent of all presumptive taxpayers in the system may be fraudulent (World Bank, 2006). Apart from such anecdotal evidence, OECD analysis shows that, while the number of small companies increased between 2000 and 2015, the total industrial output of small businesses decreased from 8.1 per cent to 5.5 per cent during that time. Along with the decrease in the average number of small business employees from eight per cent to 6.4 per cent in 2006, this evidence suggests that companies close to the simplified system thresholds of 10 or 50 employees either fragment the business or under-report employment in order to remain eligible for presumptive taxation. Despite the concentration of small businesses in booming, consumption-oriented sectors such as the retail trade, the officially reported consolidated financial results for small

[122] Based on data from 2009, when Georgia did not offer presumptive income taxation to MSMEs.

businesses in all sectors except health, education and social services was negative in 2004/05, suggesting large-scale concealment of profits (OECD, 2007).

Unusual migration trends may be an important indicator of system abuse, and the non-existence of a substantial medium-sized business category may indicate the downward migration of medium-sized businesses. Both elements can be observed in Kazakhstan, where the number of individual entrepreneurs with a registered turnover below the presumptive regime threshold of KZT 40 million suddenly increased significantly in 2009, coinciding with a reduction in the presumptive tax rate from five to three per cent.

Table 16. Kazakhstan: Number of active taxpayers by turnover in 2007–09

Turnover (million KZT)	**Legal entities***			**Individual entrepreneurs****		
	2007	**2008**	**2009**	**2007**	**2008**	**2009**
Above 2,000	1,817	1,947	1,537	1	3	5
500–2,000	4,593	4,574	3,839	47	85	57
100–500	11,765	11,757	10,448	480	757	491
40–100	10,073	10,043	9,339	820	1,120	893
20–40	8,464	8,587	8,665	2,208	2,998	4,745
Below 20	96,195	107,636	114,370	342,277	359,916	402,145
Total	**132,907**	**144,544**	**148,198**	**345,833**	**364,879**	**408,336**

Source: Tax Committee of the Ministry of Finance of the Republic of Kazakhstan, 2010

* Public agencies and associations excluded; ** taxpayers under patent regime and single land tax excluded.

The number of individual entrepreneurs above the eligibility threshold declined, and the number of incorporated businesses in the medium taxpayer segment (9,339 in 2009) was comparatively low. These might be seen as an indication of downward migration dynamics.

In a situation in which the number of businesses migrating into the presumptive regime continuously and substantially exceeds the number of businesses migrating out of it, the base of the standard tax regime erodes, tax revenue collection decreases, and competition increases for the few businesses remaining in the standard regime. As an example, migration trends in Uzbekistan from 2010 to 2012 raise concerns about the long-term effect of its presumptive tax regime.

Table 17. Uzbekistan: Number of taxpayers changing their taxation regime

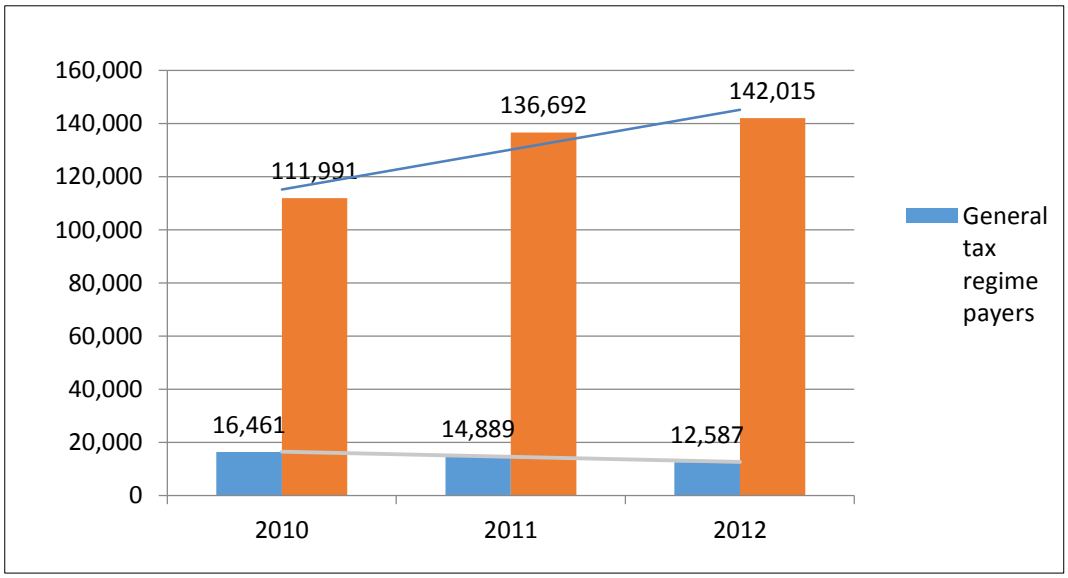
	2010	**2011**	**2012**
Number of companies that shifted from generally established tax system to simplified taxation system	1,642	1,849	1,666
Number of simplified taxpayers who shifted from simplified to generally established tax system	606	667	692

Source: Uzbekistan State Tax Committee.

In Uzbekistan, the number of net income taxpayers decreased by 25 per cent over a period of two years and is now less than 10 per cent of the taxpayer population.

The artificial splitting up of businesses in order to abuse the presumptive tax regime results in a reduction in the overall cost efficiency of business operations. Assuming rational business decision making, such splitting-up is therefore only attractive if savings in both tax payments and compliance costs exceed these efficiency losses. Eliminating differences in the tax burden between the standard and presumptive regimes substantially reduces the incentives for these business divisions. Also, the higher the efficiency losses, the smaller the newly-generated business entities need to be to qualify for the presumptive regime. The split-up option is therefore more attractive in a country like Ukraine, with a presumptive regime threshold of 50 employees and UAH 5 million ($520,000) turnover, than Latvia, with a turnover threshold of €100,000 ($113,000) and a staffing threshold of five employees. An essential step towards reducing system abuse risks is to define a presumptive regime threshold which limits regime application to small businesses facing capacity constraints and compliance difficulties with the standard tax regime.

Figure 18. Number of generally established tax payers and single tax payment payers in Uzbekistan (at 1 January each year)



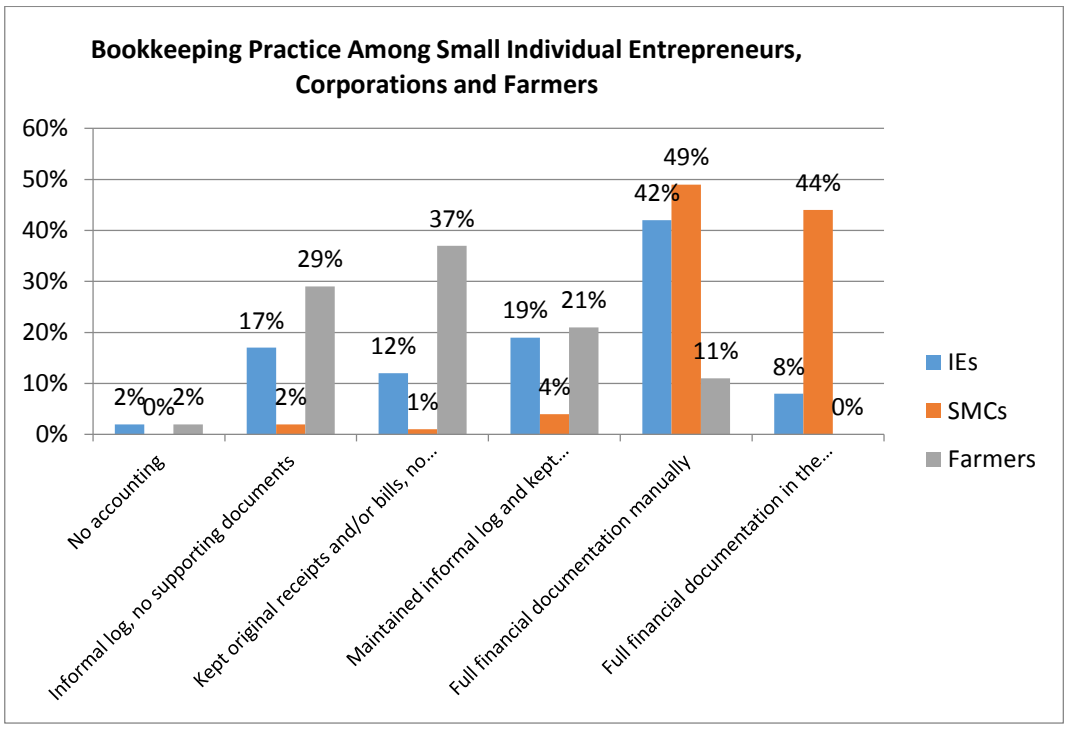
Source: State Tax Committee of Uzbekistan.

Transparent rules for the application of the regime are a second essential element. This relates in particular to the level of bookkeeping required and imposed. While a simplified cash-based bookkeeping standard is appropriate for the operation of a turnover-based presumptive tax regime, this standard is not always sufficiently enforced to verify the size of business operations in practice. For incorporated businesses, expanded reporting requirements can be considered. An extreme

approach in this respect has been taken by the Russian Federation, where an increase in the presumptive regime threshold was combined with the introduction of comprehensive accounting requirements for companies. While this is not necessarily an appropriate approach for all countries, analysis in the Kyrgyz Republic demonstrates that the enforcement of a higher level of accounting can be well in line with the actual practice of small corporations.

Due to limited transparency and control, possibilities for system abuse risks are particularly high when patent regimes with no bookkeeping requirements are extended to the small business segment. Additional safeguards need to be put in place in this case to counteract the access of larger businesses to the patent regime. In Tajikistan, for example, while the patent regime turnover threshold was relatively high at SM 200,000 ($42,000), the additional eligibility criterion of operating the business with no (non-family) employees hampered system access for larger businesses. As a result, the level of system abuse seemed rather moderate, with about 3.5 per cent of patent holders reporting, in an anonymous survey, a turnover above the patent regime threshold (SME Survey, 2009). Nevertheless, in 2013, the Tajik government used a general revision of the tax code to lower the patent threshold to SM 100,000 in order to better monitor the regime.

Figure 19: Actual bookkeeping practice of small businesses in the Kyrgyz Republic



Source: IFC SME Surveys.

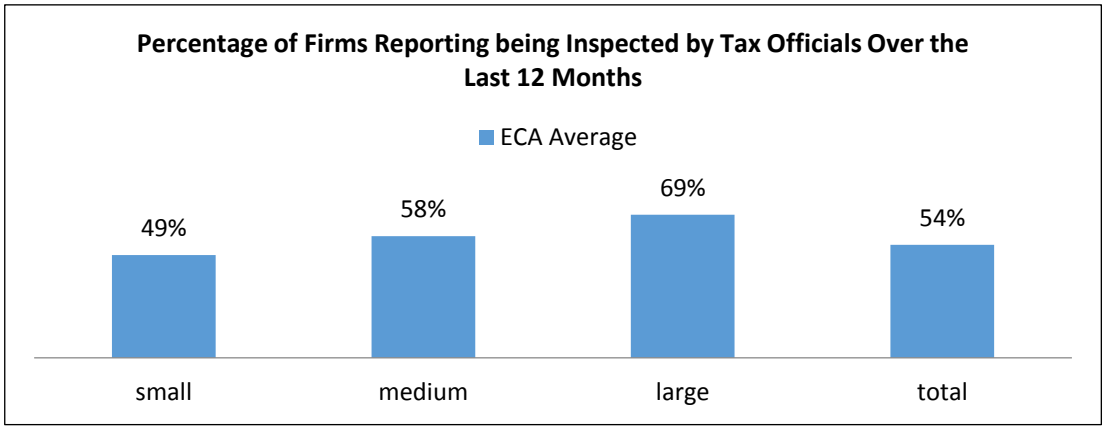
Additional system eligibility criteria and specific anti-abuse provisions create further barriers to system abuse. Such criteria may include, as in Russia, a requirement that the majority of shares (75 per cent in Russia) are owned by private individuals, that the business has no branches or representative offices, and that it

does not operate in any high-risk areas such as financial services, manufacturing of excisable goods or trading in minerals. In Latvia, in addition to restrictions regarding the number of employees, only limited liability companies having solely individuals as shareholders can apply for the micro enterprise regime.

Monitoring MSMEs

While the approach practised in the early phase of transition to target a 100 per cent audit of all registered businesses has been abolished in all countries in the region, an inappropriately high share of administrative capacity still seems to be consumed by visits to small businesses with low potential tax yields. From a business perspective, compliance costs are increased due to the time and resources required to prepare for, be available during, and respond to queries following audits.

Figure 20. Reported tax inspections in the ECA region



Source: Enterprise Survey Data, 2009-11.

The main focus of any audit programme for the small business segment must be the identification of taxpayers who abuse the presumptive tax regime for tax minimization purposes, either by substantially under-declaring their turnover or by artificially splitting up business operations. It is therefore necessary to define criteria that indicate when an artificial separation of business activities is evident. Audit activities should focus on small businesses with a turnover close to the upper threshold of the presumptive regime to verify whether, according to their actual turnover, these businesses should have migrated to the medium-sized business category. Moving to a more risk-based approach to tax audit, targeting particularly businesses that should be transferred to the general tax regime and major cases of turnover under-reporting, must therefore be a priority for tax audit reform in the region. Many Central and Eastern European countries have successfully implemented reforms in this direction, and some FSU countries, such as Kazakhstan and the Kyrgyz Republic, are following this path.

Even a well-designed and well-administered MSME regime generates only a very small and often even negligible percentage of total tax revenues.[123] This renders the administration of presumptive tax regimes relatively unattractive to national tax administrations, and several countries in the region have therefore transferred the administration of presumptive tax regimes to local governments. In such cases, successful coordination of different levels of administration is critical for effective compliance management.

In Albania, since the introduction of its first simplified scheme in the early 1990s, it has tested different special regimes and moved towards a decentralized administrative approach aiming to account for regional differences. Since 2005, income tax for small businesses with a turnover of up to ALL 8 million has been administered at the local level. For micro businesses with less than two million ALL in turnover, a simple patent applies; for businesses with a turnover between two million and eight million ALL, a turnover tax is used (with seven different turnover and three district categories).

Figure 21. Overview of presumptive taxation policies for MSMEs in Albania

Year	System/Revisions
1992	Special tax regime for individuals (trading activities, handicrafts, and other services)
1993	"Law for small business tax": fixed tax and a tax based on gross revenues (rates of 3%, 5%, and 8% on gross income)
1998	Eligibility extended to legal entities; turnover threshold of ALL 5 million introduced. Turnover tax of 4% applied to all small business with turnover between ALL 2–5 million. Fixed patent for businesses with annual turnover < ALL 2 million (differentiated by sector and location).
2002	Alignment of presumptive and VAT threshold at ALL 8 million; Fixed tax at the local level.
2005	Reduction of the rate of the simplified profits tax from 4% to 3% from 2005. Assignment of administration (and revenue) from small business tax to the local level.
2006	Turnover rate schedule introduced for 7 turnover levels and differentiated by region and sectors
2008	Introduction of balance sheet requirement for all small business with a turnover above ALL 2 million
2010	VAT threshold lowered to ALL 5 million
2014	Simplified net profit taxation with reduced rate of 7.5% for businesses with turnover between ALL 2 million and ALL 8 million. Administered by central tax administration and no longer by local governments.

The assignment of small business tax administration to local government was rather unusual, resulting in important coordination challenges between municipal authorities and regional GDT offices. Most notably, the tax base of the

[123] Tajikistan, where small businesses account for 4.2 per cent of total tax collection (patent tax 1.4 per cent, simplified tax 2.8 per cent), is an example of a comparatively high level of micro and small business contribution to total tax collection. In 2009, the Polish tax card and turnover tax regimes amounted to a total of around $500 million in tax collections, thereby accounting for just 8.5 per cent of individual business tax revenues in the country. In Armenia, following the reintroduction of the turnover tax for small businesses, the tax accounted for just 1.1 per cent of total tax revenue (EBRD, 2013)

municipalities was at risk from small business growth. Following the assignment of small business taxation to the local level, revenue dropped sharply. In Tirana, for example, between 2005 and 2010, total revenue collected from the sector dropped by almost a third, from ALL 1.4 billion to one billion, despite an increase in the number of registered small taxpayers (Tirana Municipality). There are several explanations for this decline in revenue. Political considerations resulting from conflicting agendas at the local, regional and central level are reported to have interfered with efficient tax administration. USAID (2009) highlights that the reduction in the SBT rate is the main driver of declining revenue. Also, transitional challenges and the need to build municipal authority capacity in managing small taxpayers are frequently highlighted as the main drivers of a sharp decline in revenue, particularly in light of coordination challenges. In some municipalities, rather than relying on the tax ID issued by the national registration office, a different coding system was developed to monitor the local tax base, undermining efficient information sharing. This lack of coordination has facilitated the abuse of the patent and small business regimes, as businesses that were broken up into multiple small entities and registered with different municipalities cannot be properly tracked.

Possibilities for further improvement of the presumptive regime design

One characteristic of simplified taxation in the ECA region is the continuous reform and modification of presumptive tax regimes. The directions of these reforms differ across the region, and there is thus no clear development of a new general architecture for presumptive taxation regimes. As a general observation, differences in approaches to presumptive tax design between Central and Eastern European countries and FSU countries are increasing. While, in the countries of the Former Soviet Union, a single-rate turnover tax is still the predominant approach to simplified small business taxation, CEE countries have begun to phase out pure turnover taxes and are aiming at a better alignment of presumptive and standard tax regimes.

One method of achieving such alignment is to replace turnover as the small business tax base with a simplified net income calculation. With this approach, small businesses are integrated into the standard income tax regime; however, the requirement to calculate business expenses and determine net business profit is replaced by a lump-sum cost deduction. Similar to turnover tax regimes, the business thus only has to calculate its gross income. Lump-sum cost deduction ratios may be established for different categories of small business activity to reflect average income/expense ratios. The standard income/profit tax rate is then applied on the presumed net business income.

Table 18: Examples of lump-sum cost deduction ratios in ECA

Country	Ratio
Czech Republic	80% for agricultural activities and handicraft 60% for trade 40% for any other type of activity
Slovak Republic	60% for craftsmen 25% for other businesses
Slovenia	25–70% depending on business activity

Lump-sum cost deduction elements for certain expense types are a standard feature of many personal and corporate income tax laws. Unlike a pure turnover tax, the simplified net income calculation, while facilitating compliance, integrates the business into the normal income tax regime and reduces the barrier to transition to the standard taxation regime. As in the case of turnover rate differentiation, the determination of appropriate lump-sum deduction ratios is the main challenge for system design, and lengthy discussions may emerge between ministries of finance and small business associations on acceptable deduction levels. The example of the Czech Republic illustrates this potential variability in lump-sum deduction levels.

Table 19. Changes in lump-sum expense ratios in the Czech Republic

Type of income	2004%	2005%	2010%	2011%
Income from agricultural production	50	80	80	80
Income from craft trade	25	60	80	80
Income from other trading activities	25	50	60	60
Income from an independent activity and other business categories	25	40	60	40
Income from the use or provision of industrial or other intellectual property	30	40	60	40

Source: Mincic, 2011.

The next step in the alignment of special regimes for MSMEs with the general regime would be to move to a cash flow-based net income tax for small businesses. This approach has been recommended by some experts (see, for example, Bodin & Koukpaizan, 2008, p.121) as the best possible small business taxation regime. It would probably be most suitable for countries in the ECA region – especially CEE countries – where SME operators have a comparatively high level of education and access to bookkeeping services. Such a cash-based approach is generally practised for corporate income tax in Estonia and was introduced in Hungary in 2013 with the new small business tax (KIVA) for incorporated businesses; the KIVA is a tax of 16 per cent on the adjusted sum of the company's cash-based profit and salary payments. A cash-flow tax as an alternative to a turnover tax is operated in the Russian Federation, where small businesses can opt to pay either six per cent on their turnover or 15 per cent on their net cash flow. However, the Russian approach creates the typical result associated with system competition: businesses with higher profit margins have an incentive to switch to the turnover tax regime, while

businesses with lower margins prefer the net cash flow base. For higher profit businesses, this creates a simple opportunity to reduce their tax liability (for discussion of this problem, see Alexeev & Conrad, 2013). Indeed, cash-flow based taxation of net income of small businesses is a simplification approach that avoids deviation from the general principles of income taxation.

In practice, the major problems of this approach lie in the determination and verification of deductible business expenses. In particular, a proper distinction between deductible business expenses and non-deductible private expenses and the misreporting of business expenses, which is a serious small business compliance problem even in advanced OECD countries (see, for example, GAO, 2007), may create disputes between tax administrations and businesses and increase the compliance and administrative burden. The introduction of a cash-flow tax therefore requires sufficient administrative capacity to monitor a large number of small business net income returns and extensive education of business operators in order to facilitate the correct calculation of net business income.

CONCLUSIONS

The effects of presumptive tax treatments of micro and small firms are an under-studied area in tax policy. We have presented a summary of reported effects in the ECA region, with an emerging storyline of the potential benefits in terms of compliance cost reductions, along with some indications that presumptive tax regimes encourage higher participation rates. At the same time, country experiences suggest that poorly designed small business taxation in the region acts as a disincentive for small businesses to grow to a turnover level above the presumptive regime threshold (barrier to migration upwards out of the system) and attracts larger businesses looking for areas in which to reduce their tax liability (incentive to migrate downwards into the system). Both phenomena distort business development, and thus fundamentally contradict the purpose of operating a presumptive tax regime.

The tax treatment of MSMEs differs notably across the ECA region, and the more recent move toward lump-sum cost deduction for a better alignment of presumptive and general tax treatment, seems to be an example that may find more support in future. However, the challenges of proper regime design will remain rather similar to those experienced to date. For both pure turnover taxes and lump-sum cost deductions, determination of applicable rates is often driven more by political pressures than by a sensible alignment with taxpayer profitability. Misaligned rates are problematic, given the strong incentives for high-profit activities to maintain presumptive tax treatment. When the design of tax rates and eligibility thresholds is driven by politically motivated guesswork, the risk of misalignment, which incentivizes system abuse, increases. Determining appropriate rates and defining a presumptive regime threshold that limits regime application to small businesses facing capacity constraints and difficulties in complying with the standard tax

regime therefore remains an essential step, irrespective of the simplification measures adopted.

Even the best MSME taxation design will fail to achieve its objectives of encouraging formalization and generating revenue if not implemented properly. Tailored compliance management to account for the characteristics of the segment is thus critical, and our review of the experience in ECA suggests that shortcomings in the implementation of MSME tax policy are widespread. Sometimes this is due to capacity constraints and structural challenges; sometimes it seems to be a deliberate political choice to neglect the enforcement of the rule of law for this segment; and sometimes it is a combination of the two.

REFERENCES

Alexeev, M., & Conrad, R. (2013). The Russian Tax System. In M. Alexeev, & S. Weber (Eds), *The Oxford Handbook of the Russian Economy* (pp.246-261). Oxford: Oxford University Press.

Alm, J., & Saavedra, P. (2006). *How Should Individuals be Taxed? Designing Income Taxes, Payroll Taxes, and Simplified Taxes in Ukraine*. Paper presented at the Andrew Young School of Policy Studies Annual Conference on Public Finance Issues, Georgia State University, Atlanta, GA.

Beca, M., & Serban, E. (2012). The impact of the adoption of the lump-sum tax on the SMEs' activity in Romania. *Procedia Economics and Finance, 3,* 355-360.

Bird, R. (2003). A new look at local business taxes. *Tax Notes International, 30*, 695-711.

Bird, R., & Gendron, P. (2007). *The VAT in Developing and Transition Countries*. Cambridge: Cambridge University Press.

Bird, R., & Vazquez-Caro, J. (2011). *Benchmarking Tax Administrations in Developing Countries: A Systemic Approach*. International Center for Public Policy Working Paper 85, Andrew Young School of Policy Studies, Georgia State University, Atlanta, GA.

Blažić, H., Dimitrić, M., & Škalamera-Alilović, D. (2007). Lump sum tax for small business: The case of Croatia. *Proceedings of the 14th NISPAcee Conference on Public Administration and Public Policy in Emerging Europe and Eurasia: For Professionalism, Impartiality and Transparency*, Ljubljana, Slovenia, 11-13 May.

Bodin, J., & Koukpaizan, V. (2008). Taxation of small enterprises: Recent developments. *International VAT Monitor, 2*, 121-131.

Brook, A., & Leibfritz, W. (2005). *Slovakia's Introduction of a Flat Tax as Part of Wider Economic Reforms*. OECD Economics Department Working Paper 448, OECD, Paris, France.

Bruhn, M., & Loeprick, J. (2014). *Small Business Tax Policy, Informality and Tax Evasion: Evidence from Georgia*. World Bank Policy Research Working Paper WPS 7010, World Bank, Washington, DC.

Bruna, I., & Sneidere, R. (2011). The practice of small business activities in Latvia. In N. Tsounis, A. Vlachvei, M. Gaetano, & T. Monovasilis (Eds), *Proceedings of the International Conference on Applied Economics* (pp.93-102). Department of International Trade, Kastoria Campus, Technological Institute of Western Macedonia, Greece.

Cnossen, S. (1992). Key questions in considering a value-added tax for Central and Eastern European countries. *International Monetary Fund Staff Papers, 39*(2), 211-255.

De Soto, H. (1989). *The Other Path. The Invisible Revolution in the Third World*. New York: Harper & Row.

Dickinson, P. (2012). SMEs and the business reality of Estonia's tax regulation environment. *International Journal of Law and Management*, *55*(4), 273-294.

Easterbrook, K. (2008). *Flat Taxes and Labor Supply in Central and Eastern Europe*. Stanford, CT: Stanford University.

EBRD (2013). *Research on SME: Taxation Field – Summary of the Meetings Organized in the Regions of the Republic of Armenia by the EBRD Business Support Office*. Yerevan, Armenia: European Bank for Reconstruction and Development.

Ebrill, L., Keen, M., & Summers, V. (2001). *The Modern VAT*. Washington, DC: International Monetary Fund.

Engelschalk, M. (2004). Creating a favorable environment for small business development in transition countries. In J. Alm, J. Martinez-Vazquez, & S. Wallace (Eds), *Taxing the Hard-to-Tax: Lessons from Theory and Practice* (pp.275-311). Amsterdam: Elsevier.

Engelschalk, M., & Loeprick, J. (2011). Risk scoring for SMEs. In M. Khwaja, R. Awasthi, & J. Loeprick, (Eds), *Risk-Based Tax Audits: Approaches and Country Experiences* (pp.45-54). Washington, DC: The World Bank.

European Commission (2013). *Impact Assessment: Proposal for a Council Directive Amending Directive 2006/112/EC on the Common System of Value added Tax as Regards a Standard VAT Return*. Staff Working Document SWD 427 final, European Commission, Brussels.

Fajnzylber, P., Maloney, W., & Montes-Rojas, G. (2011). Does formality improve micro-firm performance? Evidence from the Brazilian SIMPLES program. *Journal of Development Economics*, *94*(2), 262-76.

Gailute, G., Matusevicienė, L., & Buzinskiene, R. (2013). The impact of the change of the tax burden on the reliability of financial information: The case of Lithuania. *European Scientific Journal*, *9*(3), 1-18.

GAO (2007). *Tax Gap: A Strategy for Reducing the Gap Should Include Options for Addressing Sole Proprietor Noncompliance*. Report No. GAO-07-1014. Washington, DC: US Government Accountability Office.

Grabowski, M. (2011). *MSE Taxation: Polish Experiences*. Presentation for the ITD Regional Conference on SME taxation, Tbilisi, Georgia.

Holmes, K. (2002). Development of tax administrations in Central and Eastern Europe and in developing countries: Needs and opportunities for tax research. *European Taxation*, *42*(1), 18-21.

IFC (2007). *Designing a Tax System for Micro and Small Businesses: Guide for Practitioners*. Washington, DC: International Finance Corporation.

IFC (2009a). *Business Environment in Tajikistan as Seen by Small and Medium Enterprises*. Dushanbe, Tajikistan: International Finance Corporation.

IFC (2009b). *Study of Small and Medium Enterprises in Azerbaijan*. Baku, Azerbaijan: International Finance Corporation.

IFC (2009c). *The Costs of Tax Compliance in Ukraine*. Kyiv, Ukraine: International Finance Corporation.

IFC (2010). *Tax Compliance and Reporting Costs for Businesses in Uzbekistan 2010*. Tashkent, Uzbekistan: International Finance Corporation.

IFC (2011). *The Costs of Tax Compliance in Armenia.* Yerevan, Armenia: International Finance Corporation.

International Tax Dialogue (2007). *Taxation of Small and Medium Enterprises*. Background paper prepared for the ITD Conference, Buenos Aires, Argentina. Retrieved from http://www.itdweb.org/SMEconference/ConferencePapers.aspx.

International Tax Dialogue (2011). *Taxing Micro and Small Businesses: From Informality to Voluntary Tax Compliance*. Papers and presentation for a regional conference in Tbilisi, Georgia. Retrieved from http://www.itdweb.org/georgiaconference/documents/program.pdf.

Ivanova, A., Keen, M., & Klemm, A. (2005). *The Russian Flat Tax Reform*. IMF Working Paper WP/05/16, International Monetary Fund, Washington, DC.

Jack, W. (1996). *The Efficiency of VAT Implementation: A Comparative Study of Central and Eastern European Countries in Transition*. IMF Working Paper WP/96/79, International Monetary Fund, Washington, DC.

Jaworski, J. (2011a). Entrepreneurs' information needs and policy of accounting simplification in the Polish small enterprise sector: Research results, *Studia Universitatis Babes Bolyai - Negotia, 61*(4), 69-87.

Jaworksi, J. (2011b). Bookkeeping methods used in the Polish SME sector: An empirical analysis. *Zeszyty Naukowe Uniwersytetu Szczecińskiego, Finanse, Rynki Finansowe, Ubezpieczenia, 37*, 37-48.

Keen, M., Kim, Y., & Varsano, R. (2008). The "flat tax(es)": Principles and experience. *International Tax and Public Finance, 15*(6), 712-751.

Keen, M., & Mintz, J. (2004). The optimal threshold for a value-added tax. *Journal of Public Economics, 88*(3-4), 559-576.

Kireeva, E., & Rudy, K. (2006). *Tax Compliance Costs in Open Transition Economy: The Case of Belarus*. Minsk, Belarus: Belarus State Economic University.

Klun, M. (2003). Administrative costs of taxation in a transition country: The case of Slovenia. *Finance a úvěr, 53*, 75-84.

Klun, M., & Blasic, H. (2005). Tax compliance costs for companies in Slovenia and Croatia. *FinanzArchiv, 61*(3), 418-437.

Lang, M., Obermair, C., Schuch, J., Staringer, C., & Weninger, P. (2008). *Tax Compliance Costs for Companies in an Enlarged European Community*. Vienna, Austria: Linde.

Leibfritz, W. (2011). *Undeclared Economic Activity in Central and Eastern Europe*. World Bank Policy Research Working Paper WPS5923, The World Bank, Washington, DC.

Makedonskiy. S. (2005). *Taxation Mechanisms based on Simplified and Indirect Evaluation of Tax Liabilities*. Russian-European Centre for Economic Policy, Moscow, Russian Federation.

Mikhalkin, I., & Alexeeva, I. (2009). *Research into the Benefits of the Increasing Eligibility Threshold for SMEs to Participate in the Russian Simplified Tax System*. Entrepreneurship Insights e-Bulletin, Issue 8, Center for Entrepreneurship, Moscow, Russian Federation.

Mincic, L. (2011). *Overall Burden of Self-Employed Today and Tomorrow*. Paper presented at the ITD Regional Conference, Tbilisi, Georgia.

Mitra, P., and Stern, N. (2002). *Tax Systems in Transition*. World Bank Policy Research Working Paper WPS 2974, The World Bank, Washington, DC.

Monteiro, J., & Assunção, J. (2006). *Outgoing the Shadows: Estimating the Impact of Bureaucracy Simplification and Tax Cut on Formality and Investment*. Paper presented at the European Meeting of the Econometric Society, Vienna, Austria.

Morton, J. (2011). *Small Business Taxation in Eastern Europe: Experience with Presumptive Taxation 1999-2008*. Working Paper for the East and Southern Africa Universities Programme.

Nashchekina, O., & Timoshenkov. I. (2004). *Hard to Bear, Hard to Measure: The Costs of Small Business Legalization in Ukraine*. Paper presented at the International Society for New Institutional Economics Annual Conference, Tucson, AZ, 30 September-3 October.

OECD (2007). *Economic Surveys: Ukraine – Economic Assessment*. Paris, France: Organisation for Economic Co-operation and Development.

OECD (2008a). *Economic Surveys: Poland 2008*. Paris, France: Organisation for Economic Co-operation and Development.

OECD (2008b). *OECD Employment Outlook 2008*. Paris, France: Organisation for Economic Co-operation and Development.

OECD (2015). *Tax Administration 2015: Comparative Information on OECD and Other Advanced and Emerging Economies*. Paris, France: Organisation for Economic Co-operation and Development.

Packard, T., Koettl, J., & Montenegro, C. (2012). *In from the Shadow: Integrating Europe's Informal Labor*. Washington, DC: The World Bank.

Păcurari. D. (2013*). Profit Tax or Income Tax? Options for Fiscal Optimization of Romanian Small Companies*. Studies and Scientific Researches, Economics Edition No. 18, University of Bacau, Bacau, Romania.

Pashev, K. (2005). *Tax Compliance of Small Businesses in Transition Economies: Lessons from Bulgaria*. Georgia State University Working Paper 05-10, Georgia State University, Atlanta, GA.

PwC (2010). *The Impact of VAT Compliance on Business*. London: PricewaterhouseCoopers.

PwC (2013). *Study on the Feasibility and Impact of a Common EU Standard VAT Return*. London: PricewaterhouseCoopers.

PwC (2014). *Paying Taxes 2014*. London: PricewaterhouseCoopers.

Saavedra, P. (2007). Flat income tax reforms. In C. Gray, T. Lane, & A. Varoudakis (Eds), *Fiscal Policy and Economic Growth: Lessons for*

Eastern Europe and Central Asia (pp.253-280). Washington, DC: The World Bank.

Schneider, F., & Enste, D. (2000). Shadow economies: Size, causes, and consequences. *Journal of Economic Literature, 38,* 77-114.

Semikolenova, Y. (1999). *Taxation of Small and Medium Enterprises*. Thesis submitted for MA in Economics, Mohyla Academy, University of Kiev, Kiev, Russian Federation.

Semjén, Tóth, & Fazekas, M. (2008). *The Experiences of EVA in the Light of Interviews with Entrepreneurs*. Working paper, Hungarian Academy of Sciences, Budapest.

Shetinin, O., Zamulin, O., Zhuravskaya, E., & Yakovlev, E. (2005). *Monitoring the Administrative Barriers to Small Business Development in Russia: 5th Round*. Center for Economic and Financial Research Policy Paper No. 22, Moscow, Russian Federation.

Slukhai, S. (2009). Inter-location small business tax rate variation in Ukraine: What is behind it? *Financial Theory and Practice*, *33*(1), 49-71.

Stinka, J., & Bondo, D. (2014). *Micro Enterprises Tax Payers in Latvia*. SSE Riga Student Research Papers, 2014:8, Stockholm School of Economics, Stockholm, Sweden.

Tepperová, J., Pavel, J., & Láchová, L. (2015). Impact of lump sum expense rates on public revenue. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *63*(3), 1023-1030.

Thiessen, U. (2003). The impact of fiscal policy and deregulation on shadow economies in transition countries: The case of Ukraine. *Public Choice*, *114*(3), 295-318.

USAID (2009). *Collecting Taxes Database*. Retrieved from https://www.usaid.gov/data/dataset/cdeb8a1b-3440-4e88-b6cb-81b2428f8cea.

Wallace, S. (2002). *Imputed and Presumptive Taxes: International Experiences and Lessons for Russia*. Georgia State University Working Paper 02–03, Georgia State University, Atlanta, GA.

World Bank (2006). *Ukraine: Creating Fiscal Space for Growth – A Public Finance Review*. Report No. 36671-UA, The World Bank, Washington, DC.

World Bank (2011). *Tax Perception and Compliance Cost Surveys: A Tool for Tax Reform*. Washington, DC: The World Bank.