An empirical examination of the determinants of the shadow economy

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Applied Economics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>AEL-2013-0558</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>30-Sep-2013</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Acosta-González, Eduardo; Universidad de Las Palmas de Gran Canaria, Department of Quantitative Methods Fernández-Rodríguez, Fernando; Universidad de Las Palmas de Gran Canaria, Department of Quantitative Methods Sosvilla-Rivero, Simon; Universidad Complutense de Madrid, Department of Quantitative Economics</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Shadow economy, OECD countries, Tax evasion, GASIC</td>
</tr>
</tbody>
</table>
An empirical examination of the determinants of the shadow economy

Eduardo Acosta-González\textsuperscript{a}

Fernando Fernández-Rodríguez\textsuperscript{a}

Simón Sosvilla-Rivero\textsuperscript{b*}

Abstract

Using a statistical methodology guided only by data and based on a genetic algorithm, we select the best econometric model for explaining the determinants of the size of the shadow economy, its main determinants being: taxes on capital gains of individuals, corporate taxes on income, profits and capital gains, domestic credit, bank secrecy, ethnic fractionalization, urban population, globalization, corruption and the socialist legal origin of country.

JEL Codes: C51, E26, H26

\textsuperscript{a} Department of Quantitative Methods, Universidad de Las Palmas de Gran Canaria, Campus de Tafira, 35017 Las Palmas de Gran Canaria, Spain
\textsuperscript{b} Department of Quantitative Economics, Universidad Complutense de Madrid, Campus de Somosaguas, 28223 Madrid, Spain
* Corresponding author. E-mail: sosvilla@ccee.ucm.es
I. Introduction

The determinants of the shadow economy remain controversial (see Schneider and Enste, 2000, for a survey).

Using the massive database gathered by Thiessen (2010), the main contribution of our paper consists of selecting the determinants of the size of the shadow economy from a set of 274 variables suggested in the theoretical and empirical literature as potential causes and influences. We employ a purely statistical methodology which automatically selects the factors in the econometric model in a process guided only by data. Our methodology properly handles multicollinearity problems in the estimations that could arise due to the redundancy of the information provided by the factors. Besides, it offers a parsimonious model of the shadow economy, containing few factors with as much information as possible that captures the essential characteristics of the data, and can help in identifying the main reasons explaining its size.

II. Methodology

In order to avoid data mining problems for constructing the best econometric model explaining the empirical determinants of the size of the shadow economy, we follow the general to specific approach based on the theory of reduction (Hendry, 1995). It means that given an endogenous variable $Y$, which represents the indicator of the shadow economy, and a set $X_1, \ldots, X_K$ of 274 potential factors explaining it, the problem that we face is finding the best sub-model of the form:

$$ Y = \beta_0 + \beta_1 X_{i_1} + \ldots + \beta_k X_{i_k} + \varepsilon, $$

where $\{i_1, i_2, \ldots, i_k\} \subseteq \{1, 2, \ldots, K\}$, $k \leq K$, being $k$ the number of factors in the model, excluding the intercept.
The problem is that the number of sub-models where the criteria developed by econometric theory for selecting the best model is $2^K$. So, for $K = 274$, the number of possible models is higher than $10^{135}$. In order to resolve this intractable problem, we follow the heuristic strategy proposed by Hoover and Perez (1999) by searching all the feasible reduction paths when simplifying the general model. In particular, we use the methodology developed in Acosta-González and Fernández-Rodríguez (2007). This is a procedure for automatic selection of factors in the model, guided only by the data, which is carried out using a Genetic Algorithm where the lost function is the Schwarz Information Criterion (SIC, henceforth). This methodology avoids the tendency to over-identify models detected in several popular heuristic methods for selecting models [like stepwise, Lovell (1983)] and, by allowing all possible combinations of potential determinants, renders results that are robust to any specification.

Therefore, starting with a large number of potential causes and an indicator of the shadow economy considered in the literature to take into account all the information about its possible determinants, our methodology selects the best econometric model, in the sense of the SIC.

The SIC advises choosing the econometric model which minimizes the expression

$$SIC = \log \hat{\sigma}^2 + c \frac{k + 1}{N} \log(N)$$

(1)

where $\hat{\sigma}^2 = \frac{e'e}{N}$ is the variance of the residuals $e$, $N$ is the sample size, and $k$ is the number of factors in the model, excluding the intercept. The correcting factor $c$ avoids
the possibility of over-parameterized models and solves the trade-off between the in
sample goodness of fit and the out of sample forecasting ability. The higher the value of
$c$, the higher the penalty for the introduction of more factors in the model.

III. Data and empirical results

We use the database gathered by Thiessen (2010) for 38 countries for the period 1991-
2007. Given the high amount of missing data present in the Thiessen’s database, in
order to produce a data matrix without missing values, we have employed a careful case
wise deletion procedure, guided by our genetic algorithm, to eliminate the missing
observations and variables. From the initial set of 646 observations [17 years times 38
countries], the model was finally estimated with 371 observations.

Regarding the indicator of the size of the shadow economy, we use money cash holding
relative to M2. As for the explanatory variables, the database includes 294 potential
determinants of the shadow economy covering such categories as constitutional system,
tax and social security burden, administrative burden, quality of administration and
justice system, economic institutions, values and moral aspects, and other influences
and subjective factors (such as globalization, income inequality, ageing, etc.)

Table 1 reports the results provided by our genetic algorithm in order to search for the
best set of variables explaining the shadow economy. As can be seen, we obtain a

---

1 We have data on 30 OECD countries (Australia, Austria, Belgium, Canada, Cyprus,
Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Latvia,
Lithuania, Luxembourg, Malta, Mexico, Netherlands, New Zealand, Norway, Portugal,
South Korea, Spain, Sweden, Switzerland, Turkey, United Kingdom and United States)
and 8 Eastern European countries (Bulgaria, Czech Republic, Estonia, Hungary, Poland,
Romania, Slovak Republic, and Slovenia).

2 See Thiessen (2010, Appendix 1) for a detailed account of the variables included in
each of these seven major groups of potential determinants.
negative and significant association between the shadow economy and taxes on capital
gains of individuals, indicating that informality is a form of tax evasion. This is in line
with previous theoretical and empirical analyses (e.g. Constantinides 1982, Stiglitz
1983, Poterba 1987, OECD, 2006) that have suggested that there exist considerable
opportunities for avoiding tax and for generating capital losses to offset ordinary
income (especially for higher incomes), being therefore a substitute for the shadow
economy as a means of evading taxes.

We find a significant positive association between the shadow economy and taxes on
income, profits and capital gains (as a percentage of GDP), suggesting that these have a
negative impact on the decisions made by economic agents on whether to operate
officially, driving entrepreneurs and businesses underground.3

The results of the regressions in Table 1 show that domestic credit (as a percentage of
GDP) significantly influences the size of the hidden economy. This is consistent with
the fact that although domestic credit is positively and significantly correlated with
business and entry density rates, due to burdensome regulations, high marginal tax rates,
the absence of monitoring and compliance (of both registration and tax regulations), and
other weaknesses in the business environment, many firms might find it optimal to
evade regulations and operate in the informal sector (Klapper et al., 2008).

A strongly significant negative influence is also shown with regard to the bank secrecy
dummy (1 = bank secrecy law is relatively effective; 0 = no bank secrecy law or not
effective). This may indicate that bank secrecy would be a substitute for shadow

---

3 See Friedman et al. (2000) for a simple model of an entrepreneur’s decision to operate
officially or unofficially depending on corporate taxes
economy by facilitating illicit financial flows worldwide for money laundering. Indeed, the world’s biggest players in the supply of financial secrecy are rich nations (see, Walker and Unger, 2009).

The ethnic fractionalization index shows a significant negative influence on the size of the informal economy, in line with Montalvo and Reynal-Querol (2005), who document that a high degree of ethnic fractionalization has negative effects on investment and increases rent seeking activities.

With regard to urban population, our result is consistent with that of Elgin and Oyvaty (2013), who argue that although the share of the informal sector grows in the early phases of urbanization due to several pull and push factors (see, Smith, and Scarpaci, 2010), it tends to decrease in the latter phases of urbanization.

Concerning globalization, increasing competition in goods and factor markets raises uncertainty of income and employment, increasing the size of the shadow economy (Karlinger, 2009). Moreover the development of the networking-based economy has given birth to a large group of shadow enterprises, making the concept of national boundaries and distance for certain economic activities meaningless (Shangquan, 2000).

In relation to corruption, our finding supports the view that corruption and the shadow economy are complements—in high income countries and thus, a substitute for the official economy as suggested by Johnson et al. (1997) and Hindriks et al. (1999).4

4 To facilitate the interpretation, we have multiplied the original variable in Thiessen’s database by -1, so a higher value of this variable indicates greater corruption.
Finally, we detect a significant and positive influence for a dummy indicating that the legal origin of the country is socialist, suggesting that the legacy of high degree of bureaucratic discretion; undeveloped market institutions and enforcement mechanisms, and low degree of economic liberalization would have adversely influenced formal activity (Eilat and Zinnes, 2002).

Turning to the goodness of fit, the value of the R-squared statistic reported in Table 1 suggests the relative success of our regression model in predicting the shadow economy within the sample.

[Table 1]

IV.Conclusion
We have empirically examined the determinants of the size of the shadow economy using a statistical methodology guided only by data and based on a genetic algorithm. We find a positive influence for corporate taxes, domestic credit, socialist legal origin of the country and corruption, as well as a negative influence for taxes on capital gains, bank secrecy, ethnic fractionalization, urbanization and globalization.

Acknowledgements
The authors are very grateful to Ulrich Thiessen for kindly providing the dataset. Financial support from the Spanish Institute for Fiscal Studies (project IEF 13/2013) is also gratefully acknowledged.
References


Background Paper 001, Department of Economics and Social Affairs, United Nations.

overview of Vietnam and Hanoi, Urban Geography, 21, 745-757

Economics, 21, 257-294.

economic policy derived from an OECD panel analysis, International Economic
Journal, 24, 481-509.

and Economics, 2, 821–853.
TABLE 1. Estimation results for a correcting factor c=2 in the Schwarz Information Criterion

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>184.2410</td>
<td>22.7739</td>
<td>8.0900</td>
<td>0.0000</td>
</tr>
<tr>
<td>Taxes on capital gains of individuals (% of GDP)</td>
<td>-129.9080</td>
<td>10.6907</td>
<td>-12.1515</td>
<td>0.0000</td>
</tr>
<tr>
<td>Corporate taxes on income, profit &amp; c.g. (% of GDP)</td>
<td>8.2846</td>
<td>1.4892</td>
<td>5.5631</td>
<td>0.0000</td>
</tr>
<tr>
<td>Domestic credit (% of GDP)</td>
<td>1.1084</td>
<td>0.0352</td>
<td>31.5224</td>
<td>0.0000</td>
</tr>
<tr>
<td>Bank secrecy dummy)</td>
<td>-39.3740</td>
<td>5.4486</td>
<td>-7.2264</td>
<td>0.0000</td>
</tr>
<tr>
<td>Ethnic fractionalization Index</td>
<td>-174.5049</td>
<td>21.6672</td>
<td>-8.0539</td>
<td>0.0000</td>
</tr>
<tr>
<td>Urban population (% of total)</td>
<td>-1.9017</td>
<td>0.2645</td>
<td>-7.1890</td>
<td>0.0000</td>
</tr>
<tr>
<td>Globalization index</td>
<td>-1.1254</td>
<td>0.2939</td>
<td>-3.8294</td>
<td>0.0002</td>
</tr>
<tr>
<td>Corruption</td>
<td>8.3117</td>
<td>2.0670</td>
<td>4.0206</td>
<td>0.0001</td>
</tr>
<tr>
<td>Legal origin of country, Socialistic</td>
<td>38.4636</td>
<td>8.4153</td>
<td>4.5707</td>
<td>0.0000</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.8119</td>
<td>F-statistic</td>
<td>173.1621</td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.8073</td>
<td>Prob(F-statistic)</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>Standard Error of regression</td>
<td>39.5097</td>
<td>Schwarz criterion</td>
<td>24.1386</td>
<td></td>
</tr>
</tbody>
</table>