The causal relationship between unemployment rate and U.S. shadow economy.
A Toda-Yamamoto approach

Adriana AnaMaria Alexandru, Ion Dobre and Catalin Corneliu Ghinararu

Abstract—The paper analyses the causal relationship between U.S. shadow economy (SE) and unemployment rate (UR) using Toda-Yamamoto approach for quarterly data covering the period 1980-2009. The size of the shadow economy as % of official GDP is estimated using a MIMIC model with four causal variables (taxes on corporate income, contributions for government social insurance, unemployment rate and self-employment) and two indicators (index of real GDP and civilian labour force participation rate). Their dimension is decreasing over the last two periods.

The evidence generally supports the existence of a unidirectional causality that runs from unemployment rate to shadow economy for the case of United States.

Keywords—shadow economy, unemployment rate, MIMIC model, Toda-Yamamoto approach, USA.

I. INTRODUCTION

The relationship between the shadow economy (SE) and the level of unemployment is one of major interest. People work in the shadow economy because of the increased cost that firms in the formal sector have to pay to hire a worker. The increased cost comes from the tax burden and government regulations on economic activities. In discussing the growth of the shadow economy, the empirical evidence suggests two important factors: (a) reduction in official working hours, (b) the influence of the unemployment rate (UR).

Enste [20] points out that the reduction of the number of working hours below worker's preferences raises the quantity of hours worked in the shadow economy. Early retirement also increases the quantity of hours worked in the shadow economy.

In Italy, Bertola and Garibaldi [6] present the case that an increase in payroll taxation can have effect on the supply of labour and the size of the shadow economy. An increase in tax and social security burdens not only reduces official employment but tends to increase the shadow labour force. This is because an increase in payroll tax can influence the decision to participate in official employment.

Also, Boeri and Garibaldi [8] show a strong positive correlation between average unemployment rate and average shadow employment across 20 Italian regions during the period 1995-1999.

Dell’Anno and Solomon [11] find a positive relationship between unemployment rate and shadow economy using a SVAR analysis, showing that a positive aggregate supply shock will cause in increase in the shadow economy by about 8% above the baseline.

The paper analyzes the causal relationship between shadow economy and unemployment rate using Toda-Yamamoto approach.

II. DATA AND METHODOLOGY

II.1. DATA ISSUES

The data series used in the study are quarterly, seasonally adjusted covering the period 1980:Q1 to 2009:Q2. The main source of data is U.S. Bureau of Economic Analysis, U.S. Bureau of Labour Statistics Data and Federal Reserve Bank.

The series in levels or differences have been tested for unit roots using the Augmented Dickey Fuller (ADF) test and PP tests. All the data has been differentiated for the achievement of the stationarity. While all the variables have been identified like integrated on first order, the latent variable is estimated in the same transformation of independent variables (first difference).

II.2. METHODOLOGY

The size of the U.S. shadow economy is estimated as % of official GDP using a particular type of structural equations models-MIMIC model.

The MIMIC model- Multiple Indicators and Multiple Causes model (MIMIC model), allows to consider the SE as a “latent” variable linked, on the one hand, to a number of observable indicators (reflecting changes in the size of the SE) and on the other, to a set of observed causal variables, which are regarded as some of the most important determinants of the unreported economic activity [8].
The model is composed by two sorts of equations, the structural one and the measurement equations system. The equation that captures the relationships among the latent variable ($\eta$) and the causes ($X'$) is named “structural model” and the equations that links indicators ($Y'$) with the latent variable (non-observed economy) is called the “measurement model”. A MIMIC model of the hidden economy is formulated mathematically as follows:

$$ Y = \lambda \eta + \epsilon $$

$$ \eta = \gamma X + \xi $$

where:
- $\eta$ is the scalar latent variable (the size of shadow economy);
- $Y' = (Y_1, ..., Y_p)$ is the vector of indicators of the latent variable;
- $X' = (X_1, ..., X_q)$ is the vector of causes of $\eta$;
- $\lambda_{(p \times 1)}$ and $\gamma_{(q \times 1)}$ vectors of parameters;
- $\epsilon_{(p \times 1)}$ and $\xi_{(q \times 1)}$ vectors of scalar random errors;

The $\epsilon'$s and $\xi$ are assumed to be mutually uncorrelated. Substituting (2) into (1), the MIMIC model can be written as:

$$ Y = \Pi X + z $$

where: $\Pi = \lambda \gamma$, $z = \lambda \xi + \epsilon$.

The estimation of (1) and (2) requires a normalization of the parameters in (1), and a convenient way to achieve this is to constrain one element of $\lambda$ to some pre-assigned value ([21]-[22]).

The possible causes of shadow economy considered in the model are: tax burden decomposed into personal current taxes ($X_1$), taxes on production and imports ($X_2$), taxes on corporate income ($X_3$), contributions for government social insurance ($X_4$) and government unemployment insurance ($X_5$), unemployment rate ($X_6$), self-employment in civilian labour force ($X_7$), government employment in civilian labour force ($X_8$) called bureaucracy index. The indicator variables incorporated in the model are: real gross domestic product index ($Y_1$), currency ratio $M_1/M_2$ ($Y_2$) and civilian labour force participation rate ($Y_3$).

The variables used into the estimation of the shadow economy are also quarterly and seasonally adjusted covering the period 1980-2009. All the data has been differentiated for the achievement of the stationarity.

In order to estimate the MIMIC model, by Maximum Likelihood, using the LISREL 8.8 package, we normalized the coefficient of the index of real GDP ($\lambda_1 = -1$) to sufficiently identify the model. This indicates an inverse relationship between the official and shadow economy.

In order to identify the best model, we have started with MIMIC model 8-1-3 and we have removed the variables which have not structural parameters statistically significant.

A detailed description and implementation of the MIMIC model for the USA shadow economy is provided in [15].

After we estimate the size of the shadow economy, we investigate the nature of the relationship between the two variables using Toda-Yamamoto approach.

Toda and Yamamoto [33] causality test is applied in level VARs irrespective of whether the variables are integrated, cointegrated, or not. Toda and Yamamoto [33] argue that F-statistic used to test for traditional Granger causality may not be valid as the test does not have a standard distribution when the time series data integrated or cointegrated.

The Toda-Yamamoto procedure basically involves estimation of an augmented VAR ($k + d_{max}$) model, where $k$ is the optimal lag length in the original VAR system and $d_{max}$ is maximal order of integration of the variables in the VAR system.

The Toda-Yamamoto causality test applies a modified Wald (MWALD) test statistic to test zero restrictions on the parameters of the original VAR ($k$) model. The test has an asymptotic (chi-square) distribution with $k$ degrees of freedom. The test essentially involves two stages. The first stage determines the optimal lag length ($k$) and the maximum order of integration ($d$) of the variables in the system. The lag length, $k$ is obtained in the process of the VAR in levels among the variables in the system by using different lag length criterion such as AIC or SBC. The unit root testing procedure, such as Dickey-Fuller [12] ADF and Phillips-Perron [29] tests may be used to identify the order of integration, $d$.

The second stage uses the modified Wald procedure to test the VAR ($k$) model for causality. The optimal lag length is equal to $p = [k + d_{max}]$. In the case of a bivariate ($Y, X$) relationship, Toda and Yamamoto [33] causality test is represented as follows:

$$ Y_t = a_0 + \sum_{i=1}^{k} a_i Y_{t-i} + \sum_{i=k+1}^{k+d_{max}} a_{i-k} Y_{t-i} + \sum_{i=1}^{k} b_i X_{t-i} + \sum_{i=k+1}^{k+d_{max}} b_{i-k} X_{t-i} + \epsilon_{1t} $$

$$ X_t = d_0 + \sum_{i=1}^{k} d_i Y_{t-i} + \sum_{i=k+1}^{k+d_{max}} d_{i-k} Y_{t-i} + \sum_{i=1}^{k} e_i X_{t-i} + \sum_{i=k+1}^{k+d_{max}} e_{i-k} X_{t-i} + \epsilon_{2t} $$

The Wald tests were then applied to the first $k$ coefficients matrices using the standard $\chi^2$ statistics (Duasa[17]). Let $c_i = vec(c_{i1}, c_{i2}, ..., c_{ik})$ be the vector of the first $k$ VAR coefficients.

The null hypothesis that $X$ does not cause $Y$ is constructed as follows: $H_0 : c_{it} = 0, \ i = 1, ..., k$.

Similarly the second null hypothesis that $Y$ does not cause $X$ is formulated as follows: $H_0 : \vec{e}_{it} = 0, \ i = 1, ..., k$. The system given by equations (4)-(5) is estimated using the
Seemingly Unrelated Regression technique (Rambaldi and Doran[30]). A Wald test is then carried out to test the hypothesis. The computed Wald-statistic has an asymptotic chi-square distribution with k degrees of freedom.

III. EMPIRICAL RESULTS

III.1. ESTIMATING THE SIZE OF SHADOW ECONOMY

In order to estimate the size of the shadow economy, we have identified the best model as MIMIC 4-1-2 with four causal variables (taxes on corporate income, contributions for government social insurance, unemployment rate and self-employment) and two indicators (index of real GDP and civilian labour force participation rate).

Taking into account the reference variable \( Y_{1990} \) the shadow economy is scaled up to a value in 1990, the base year, and we build an average of several estimates from this year for the U.S.A. shadow economy (table 1).

The index of changes of the shadow economy \( \eta \) in United States measured as percentage of GDP in the 1990 is linked to the index of changes of real GDP as follow:

Measurement Equation:

\[
\frac{GDP_t - GDP_{t-1}}{GDP_{1990}} = \frac{\tilde{\eta}_t - \tilde{\eta}_{t-1}}{GDP_{1990}}
\]

I. Estimates of the size of U.S.A. shadow economy (1990)

<table>
<thead>
<tr>
<th>Author</th>
<th>Method</th>
<th>Size of Shadow Economy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lacko(1999)</td>
<td>Physical Input(Electricity)</td>
<td>10.5%</td>
</tr>
<tr>
<td>Schneider and Enste(2000)</td>
<td>Currency Demand Approach</td>
<td>7.5%*</td>
</tr>
</tbody>
</table>

*means for 1990-1993

The estimates of the structural model are used to obtain an ordinal time series index for latent variable (shadow economy):

Structural Equation:

\[
\frac{\Delta \tilde{\eta}}{GDP_{1990}} = -0.24X_3 + 3.00X_4 + 1.49X_5 + 1.01X_7
\]

The index is scaled to take up to a value of 10.6% in 1990 and further transformed from changes respect to the GDP in the 1990 to the shadow economy as ratio of current GDP:

\[
\frac{\tilde{\eta}}{GDP_{1990}} \times \frac{\eta^*}{GDP_{1990}} \times \frac{GDP_{1990}}{\tilde{\eta}_{1990}} \times \frac{GDP_{1990}}{GDP_t} = \frac{\tilde{\eta}_t}{GDP_t}
\]

I. \( \frac{\tilde{\eta}_{1990}}{GDP_{1990}} \) is the index of shadow economy calculated by (7);

II. \( \frac{\eta^*}{GDP_{1990}} = 10.6\% \) is the exogenous estimate of shadow economy;

III. \( \frac{\tilde{\eta}_{1990}}{GDP_{1990}} \) is the value of index estimated by (7);

IV. \( \frac{GDP_{1990}}{GDP_t} \) is to convert the index of changes respect to base year in shadow economy respect to current GDP;

V. \( \frac{\tilde{\eta}_t}{GDP_t} \) is the estimated shadow economy as a percentage of official GDP.

The shadow economy measured as percentage of official GDP records the value of 13.41% in the first trimester of 1980 and follows an ascendant trend reaching the value of 16.77% in the last trimester of 1982. At the beginning of 1983, the dimension of USA shadow economy begins to decrease in intensity, recording the average value of 6% of GDP at the end of 2007. For the last two year 2008 and 2009, the size of the unreported economy it increases slowly, achieving the value of 7.3% in the second quarter of 2009.

The results are not far from the last empirical studies for USA ([20], [31], [32]). Schneider estimates in his last study, the size of USA shadow economy as % of GDP, at the level of 7.9% in 2005, respectively 8% in 2006.

III.2. THE RELATIONSHIP BETWEEN UNEMPLOYMENT RATE AND U.S. SHADOW ECONOMY

In many empirical studies, it has been found that tax burden is the biggest causes of shadow economy. Also the size of shadow economy is influenced by the level of unemployment. An increase in unemployment rates reduces the proportion of workers employed in the formal sector, this leads to higher labor participation rates in the informal sector.
The graphical evolution of the shadow economy versus unemployment rate reveal the existence of a strong positive relationship between the two variables, quantified by a value of about 0.80 of correlation coefficient.

Fig. 2. Shadow economy vs. Unemployment rate in United States

Giles ([21], [22]) states that the effect of unemployment on the shadow economy is ambiguous (i.e. both positive and negative). An increase in the number of unemployed increases the number of people who work in the black economy because they have more time. On the other hand, an increase in unemployment implies a decrease in the shadow economy. Thus the null hypothesis which requires finding the significance of the Wald test is to see if the coefficients of the lagged “other” variables (excluding the additional one) are jointly zero in the equation (Duasa [17]). To test that UR does not Granger cause SE, we estimate the VAR (3) model and test that $UR_{t-1}, UR_{t-2}$ does not appear in SE equation. Thus the null hypothesis is $H_0: a_{11}^{(1)} = a_{12}^{(2)} = 0$ where $a_{ij}^{(l)}$ are the coefficients of $UR_{t-i}, i = 1,2$ in the first equation of the system.

The existence of causality from unemployment rate to shadow economy can be established through rejecting the above null hypothesis which requires finding the significance of the MWald statistic for the group of the lagged independent variables identified above.

II. ADF and PP tests for Unit Root analysis

<table>
<thead>
<tr>
<th></th>
<th>Shadow Economy(SE)</th>
<th>Unemployment rate(UR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T&amp;C</td>
<td>C</td>
</tr>
<tr>
<td>Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADF</td>
<td>-1.08</td>
<td>-1.39</td>
</tr>
<tr>
<td>lag</td>
<td>(5)</td>
<td>(5)</td>
</tr>
<tr>
<td>PP</td>
<td>-1.26</td>
<td>-0.92</td>
</tr>
<tr>
<td>lag</td>
<td>(6)</td>
<td>(6)</td>
</tr>
<tr>
<td>First diff</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADF</td>
<td>-3.43*</td>
<td>-3.39***</td>
</tr>
<tr>
<td>lag</td>
<td>(2)</td>
<td>(2)</td>
</tr>
<tr>
<td>PP</td>
<td>-5.99*</td>
<td>-5.97*</td>
</tr>
<tr>
<td>lag</td>
<td>(5)</td>
<td>(5)</td>
</tr>
</tbody>
</table>

Note: T&C represents the most general model with a drift and trend; C is the model with a drift and without trend; None is the most restricted model without a drift and trend. Numbers in brackets are lag lengths used in ADF test (as determined by SCH set to maximum 12) to remove serial correlation in the residuals. When using PP test, numbers in brackets represent Newey-West Bandwith (as determined by Bartlett-Kernel). Both in ADF and PP tests, unit root tests were performed from the most general to the least specific model by eliminating trend and intercept across the models (See Enders, 1995: 254-255). *, ** and *** denote rejection of the null hypothesis at the 1%, 5% and 10% levels respectively. Tests for unit roots have been carried out in E-VIEWS 6.0.

Since $d_{max} = 1$, we must estimate a VAR (3) for the relationship between unemployment rate and shadow economy:

$$x_t = A_0 + A_1 x_{t-1} + A_2 x_{t-2} + A_3 x_{t-3} + A_4 x_{t-4} + \varepsilon_t$$  \hspace{1cm} (9)

$$\begin{bmatrix} SE \\ UR \end{bmatrix} = \begin{bmatrix} a_{10}^{(0)} & a_{11}^{(0)} & d_{11}^{(0)} \\ a_{20}^{(0)} & a_{21}^{(0)} & d_{21}^{(0)} \\ SE_{t-1} & SE_{t-2} & SE_{t-3} \end{bmatrix} \begin{bmatrix} d_{11}^{(1)} & d_{12}^{(1)} \\ d_{21}^{(1)} & d_{22}^{(1)} \end{bmatrix} \begin{bmatrix} SE_{t-1} \\ SE_{t-2} \\ SE_{t-3} \end{bmatrix} + \begin{bmatrix} e_{t1} \\ e_{t2} \end{bmatrix}$$ \hspace{1cm} (10)

where:

$$E(e_t) = \begin{bmatrix} e_{t1} \\ e_{t2} \end{bmatrix} = 0 \quad \text{and} \quad E(e_t e_t') = \Sigma.$$
III. The results of the Toda-Yamamoto causality test

<table>
<thead>
<tr>
<th>Null hypothesis</th>
<th>p</th>
<th>MWald statistics</th>
<th>p-values</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0 : GR$ does not Granger cause SE</td>
<td>3</td>
<td>12.06</td>
<td>0.0024*</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td>$H_0 : SE$ does not Granger cause UR</td>
<td>3</td>
<td>0.91</td>
<td>0.955</td>
<td>Do not reject $H_0$</td>
</tr>
</tbody>
</table>

* * indicates rejection of the null at the 1% level, respectively 5% level

According to the Toda-Yamamoto causality test results shown in Table III, there is strong evidence of causality running from unemployment rate to shadow economy at the 1% level of significance. The results do not reveal causality from shadow economy to unemployment rate. Therefore, we can conclude that there is a uni-directional direction of causality that runs from unemployment rate to shadow economy for the case of United States.

IV. CONCLUSIONS

The paper has investigated the nature of the relationship between unemployment rate and the size of the U.S.A. shadow economy measured as % of official GDP for the period 1980-2009, using Toda-Yamamoto approach. The size of the shadow economy estimated using the MIMIC model is decreasing over the last two periods, achieving the value of about 7.3% of official GDP at the middle of 2009.

The empirical results point out that there is strong evidence of uni-directional causality running from unemployment rate to shadow economy at the 1% level of significance.

REFERENCES


*** Eviews 6.0 software
*** Lisrel 8.8 package