Revisiting the relationship between unemployment rate and the size of the shadow economy for United States using Johansen approach for cointegration

ADRIANA ANAMARIA ALEXANDRU  
National Scientific Institute for Labour and Social Protection  
6-8 Povernei Street, Bucharest, ROMANIA  
Statistics and Econometrics Department  
Academy of Economic Studies Bucharest  
15-17, Calea Dorobantilor, Bucharest, ROMANIA  
adrianaalexandru@yahoo.com

ION DOBRE  
Economic Cybernetics Department  
Academy of Economic Studies Bucharest  
15-17, Calea Dorobantilor, Bucharest, ROMANIA  
dobrerio@ase.ro

CATALIN CORNELIU GHINARARU  
National Scientific Institute for Labour and Social Protection  
6-8 Povernei Street, Bucharest, ROMANIA  
ghinararu@incsmps.ro

Abstract: The paper revisits and investigates the possible co-integration and direction of causality between unemployment rate(UR) and the size of the shadow economy(SE) in case of USA, by employing Johansen and Granger approaches, using quarterly data covering the period 1980-2009. The shadow economy is estimated as percentage of official GDP, using MIMIC model and their dimension is decreasing over the last two decades. The empirical results reveal the existence of a long run relationship between the two variables. Furthermore, the Granger causality tests identify a unidirectional causality that runs from unemployment rate to the size of shadow economy.

Keywords: shadow economy, unemployment rate, MIMIC model, cointegration, Johansen approach, VECM, Granger causality, USA.

1. Introduction

The relationship between the shadow economy 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.

Enste [12] 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 [1] 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 [2] show a strong positive correlation between average unemployment rate and average shadow employment across 20 Italian regions between 1995-1999.
The paper examines the possible co-integration relationship between unemployment rate and the size of the shadow economy and tests the direction of causality between these variables.

2. Data and Methodology

The study used quarterly data covering the period 1980-2009. 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 [4].

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 [10].

After we estimate the size of the shadow economy, we investigate the nature of the relationship between the two variables.

The Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) Unit Root Tests are employed to test the integration level and the possible co-integration among the size of the shadow economy estimated using MIMIC model and the unemployment rate ([7], [22]).

After the order of integration is determined, co-integration between the series should be tested to identify any long run relationship. Johansen trace test is used for the co-integration test in this study. Cheung and Lai [3] mention that the trace test is more robust than the maximum eigenvalue test for co integration. The Johansen trace test attempts to determine the number of co-integrating vectors among variables. There should be at least one co-integrating vector for possible co-integration. This procedure \(^{1}[19]\) can be expressed in the following VAR model:

\[
X_t = \Pi_1 X_{t-1} + \ldots + \Pi_K X_{t-K} + \mu + e_t \quad t = 1, \ldots, T \tag{1}
\]

where \(X_t, X_{t-1}, \ldots, X_{t-K}\) are vectors of current and lagged values of \(P\) variables which are \(I(I)\) in the model; \(\Pi_1, \ldots, \Pi_K\) are matrices of coefficients with \((P \times P)\) dimensions; \(\mu\) is an intercept vector; and \(e_t\) is a vector of random errors. The number of lagged values, in practice, is determined in such a way that error terms are not significantly auto-correlated. The rank of \(\Pi\) is the number of co-integrating relationship(s) (i.e. \(r\)) which is determined by testing whether its Eigen values \((\lambda_{r})\) are statistically different from zero. Johansen and Juselius [15] propose that using the Eigen values of \(\Pi\) ordered from the largest to the smallest is for computation of trace statistics\(^2\). The trace statistic \((\Lambda_{trace})\) is computed by the following formula\(^3\):
\[ \lambda_{trace} = -T \sum \ln(1 - \lambda_i) \]  
\[ i = r+1, \ldots, n-1 \] 
and the hypotheses are:

- **H₀**: \( r = 0 \)
- **H₁**: \( r \geq 1 \)
- **H₀**: \( r \leq 2 \)
- **H₁**: \( r \geq 3 \)

If the series are I(1) and cointegrated, then Granger Causality tests should be run under VECM framework([20], [19]):

\[ \Delta Y_t = C_0 + \sum_{i=1}^{k} \beta_i Y_{t-i} + \sum_{i=1}^{k} \alpha_i X_{t-i} + p_i ECT_{t-i} + u_t \]  
\[ \Delta X_t = C_0 + \sum_{i=1}^{k} \gamma_i X_{t-i} + \sum_{i=1}^{k} \zeta_i Y_{t-i} + \eta_i ECT_{t-i} + \epsilon_i \]

Where \( Y, X \) are the variables, \( \alpha_i, p_i \) are the adjustment coefficient while \( ECT_{t-i} \) expresses the error correction term. In eq.(3), \( X \) Granger causes \( Y \) if \( \alpha_i, p_i \) are significantly different from zero. In eq.(4) \( Y \) Granger causes \( X \) if \( \zeta_i, \eta_i \) are significantly different from zero. F-test alone is not enough to have causation; t-ratio of ECM term should be also negative and statistically significant together with F value of the model to have causation in the models.

### 3. Empirical results

#### 3.1. Estimating the size of the 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 \( Y_{1990} = \text{Real GDP}_{1990} \) 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:

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

### Table 1: 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>
<tr>
<td><strong>Mean 1990</strong></td>
<td></td>
<td><strong>10.6%</strong></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:**

\[ \frac{\Delta \tilde{\eta}_t}{GDP_{1990}} = -0.24X_{1990} + 3.04X_{1990} + 1.49X_{1990} + 1.04X_{1990} \]

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:

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

- I. \( \frac{\tilde{\eta}_t}{GDP_{1990}} \) is the index of shadow economy calculated by eq.(6);
- II. \( \frac{\eta^*_{1990}}{GDP_{1990}} = 10.6% \) is the exogenous estimate of shadow economy;
- III. \( \frac{\tilde{\eta}_{1990}}{GDP_{1990}} \) is the value of index estimated by eq.(6);
- IV. \( \frac{GDP_{1990}}{GDP_{1990}} \) is to convert the index of changes respect to base year in shadow economy respect to current GDP;
- V. \( \frac{\hat{\eta}_t}{GDP_{1990}} \) is the estimated shadow economy as a percentage of official GDP.
The shadow economy measured as a 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 years 2008 and 2009, the size of the unreported economy 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 ([12], [24], [25]). Schneider estimates in his last study, the size of USA shadow economy as average 2004/05, at the level of 7.9 percentage of official GDP.

### 3.2. Evaluation of the relationship between the shadow economy and the unemployment rate under Johansen approach

The main goal of the study is to investigate the nature of the relationship between the two variables and to identify any possible direction of causality between them. In order to identify the level of integration of the two series, ADF and PP unit root tests were applied; the results are presented in table 2. The size of the shadow economy seems to be stationary in ADF test at level but this is not justified by PP test. Further more, both tests reveal that the variables are non-stationary at their levels but stationary at their first differences, being integrated of order one, I(1).

#### Table 2. ADF and PP Tests for Unit Root

<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>-3.09</td>
<td>-1.39</td>
</tr>
<tr>
<td>lag</td>
<td>(3)</td>
<td>(3)</td>
</tr>
<tr>
<td>PP</td>
<td>-2.26</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>-6.99*</td>
<td>-6.97*</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.

Because the both series are integrated of the same order, I(1) we will apply Johansen and Juselius[15] cointegration approach in order to investigate if there is a long run relationship between the two variables.

Although the optimal number of lags is two, established by SIC and HQ criterions, Pindyck and Rubinfeld [25] pointed out that it would be best to run the test for a few different lag structures and make sure that the results were not sensitive to the choice of lag length.

In table 2 are presented the results of co-integration tests using Johansen and Juselius approach(1990) and confirms that there is a unique co-integration vector(a long run relationship) between the two variables.

According to the normalized parameter estimates, we can conclude that unemployment rate has a positive and elastic effect on the size of the shadow economy. When unemployment rate grows by 1% the U.S. shadow economy will rise with about 2.34%.

Because a long run equilibrium relationship is found between unemployment rate and the size of the shadow economy, a VECM model is constructed to determine the direction of causality. Table 4 reports the F-statistics and t-statistics for error correction term defined for the null hypothesis of no-causality.
Table 3. Cointegration tests using the Johansen (1988) and Johansen and Juselius (1990) approach

<table>
<thead>
<tr>
<th>Variables</th>
<th>Trace statistic</th>
<th>5% Critical Value</th>
<th>1% Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lag 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UR, SE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_0: r = 0$</td>
<td>25.41**</td>
<td>12.53</td>
<td>16.31</td>
</tr>
<tr>
<td>$H_1: r \leq 1$</td>
<td>0.70</td>
<td>3.84</td>
<td>6.51</td>
</tr>
<tr>
<td>Lag 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UR, SE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_0: r = 0$</td>
<td>21.00**</td>
<td>12.53</td>
<td>16.31</td>
</tr>
<tr>
<td>$H_1: r \leq 1$</td>
<td>0.14</td>
<td>3.84</td>
<td>6.51</td>
</tr>
<tr>
<td>Lag 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UR, SE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_0: r = 0$</td>
<td>13.31*</td>
<td>12.53</td>
<td>16.31</td>
</tr>
<tr>
<td>$H_1: r \leq 1$</td>
<td>0.04</td>
<td>3.84</td>
<td>6.51</td>
</tr>
<tr>
<td>Lag 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UR, SE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_0: r = 0$</td>
<td>7.42</td>
<td>12.53</td>
<td>16.31</td>
</tr>
<tr>
<td>$H_1: r \leq 1$</td>
<td>0.06</td>
<td>3.84</td>
<td>6.51</td>
</tr>
</tbody>
</table>

Note:
Trace test indicates 1 co-integrating equation(s) at both 5% and 1% levels for lag 1 and 2, and 1 cointegrating equation at 5% level. **(** denotes rejection of the hypothesis at the 5% (1%) level.

Because the t-ratio of ECT is positive and not statistically significant, we can conclude that we don’t have any Granger causality from SE to UR, but we can say that we have a unidirectional causality from UR to SE (t-ratio of ECT and F-ratio are statistically significant at 1% and 5% levels, but the ECT is not negative).

Table 4. Granger Causality Tests

<table>
<thead>
<tr>
<th>Null hypothesis</th>
<th>UR does not Granger cause SE</th>
<th>SE does not Granger cause UR</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-stat</td>
<td>22.42*</td>
<td>39.37*</td>
</tr>
<tr>
<td>Lag 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{ECT,-1}$</td>
<td>2.63**</td>
<td>1.47</td>
</tr>
<tr>
<td>F-stat</td>
<td>12.94*</td>
<td>25.96*</td>
</tr>
<tr>
<td>Lag 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{ECT,-1}$</td>
<td>2.40**</td>
<td>2.062</td>
</tr>
<tr>
<td>F-stat</td>
<td>11.14*</td>
<td>19.99*</td>
</tr>
<tr>
<td>Lag 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{ECT,-1}$</td>
<td>2.50**</td>
<td>1.78</td>
</tr>
</tbody>
</table>

* and ** denote significance for 1% and 5% levels.

Table 5. Estimation of the Granger Causality Tests within Block Exogeneity Wald Tests

<table>
<thead>
<tr>
<th>Dependent variable: SE</th>
<th>$X^2$</th>
<th>Lag 1</th>
<th>Lag 2</th>
<th>Lag 3</th>
<th>Lag 4</th>
<th>Lag 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclude</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UR</td>
<td>0.06</td>
<td>0.02</td>
<td>0.37</td>
<td>4.48</td>
<td>7.66</td>
<td></td>
</tr>
</tbody>
</table>

* and ** denote significance for 1% and 5% levels.

4. Conclusions

The paper has investigated the existence of a long run relationship and direction of any causality between unemployment rate and the size of the U.S.A. shadow economy measured as % of official GDP for the period 1980-2009. The size of the shadow economy estimated using the MIMIC model is decreasing over the last two decades, from thirteen to seventeen percent between 1980 and 1983 up to 7 percent of official GDP in 2009.

The empirical results point out the existence of a unique co-integration relationship between the variables. Thus, there is a unidirectional causation that runs from unemployment rate to shadow economy.

Acknowledgement

The author Alexandru Adriana would like to thank Professor Salih Turan Katircioglu (Department of Banking and Finance, Eastern Mediterranean University, North Cyprus) for valuable knowledge imparted during “The Fourth International Training Workshop on Econometric Modeling and Data Analysis “.

References


*** Eviews 6.0 software

*** Lisrel 8.8 package