The relationship between unemployment rate and the size of the shadow economy- A nonparametric analysis of U.S. data with spline models

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

Abstract: The paper aims to investigate the relationship between unemployment rate and shadow economy with USA data using spline models. The shadow economy is estimated as percentage of official GDP, using MIMIC model. The size of the shadow economy (SE) is estimated to be decreasing over the last two decades.

In order to evaluate the nature of the relationship between the two variables, we have estimated cubic B-spline, natural cubic B-spline and smoothing models.Using an F-test, we compare the smoothing spline to a global linear fit and the results indicate a sufficiently linear relationship.

Finally, we have compared the local polynomial models with the spline model; the smoothing spline model closely matches the linearity between the size of the shadow economy and the unemployment rate.

Keywords: shadow economy, unemployment rate, MIMIC model, spline models, 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 [10] 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 [2] 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 [3] 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 relationship between unemployment rate and the size of the shadow economy using a nonparametric analysis based on spline models.

2. Estimating the size of the U.S. shadow economy

2.1. 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. The main sources of the data are U.S. Bureau of Economic Analysis (BEA) and U.S. Bureau of Labor Statistics (BLS).

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 [9].

2.2. Empirical results

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 ($Y_1, \frac{Real\,GDP_t}{Real\,GDP_{1990}}$) 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 (η) 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: $$\frac{GDP_t - GDP_{t-1}}{GDP_{1990}} = \frac{\tilde{\eta}_t - \tilde{\eta}_{t-1}}{GDP_{1990}} \quad (1)$$

Table 1: Estimates of the size of U.S.A. shadow economy (1990)

Author	*Method*	*Size of Shadow Economy*
Johnson et. Al(1998)	Currency Demand Approach	13.9%
Lacko(1999)	Physical Input(Electricity)	10.5%
Schneider and Enste(2000)	Currency Demand Approach	7.5%*
Mean 1990		***10.6%***

*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:

$$\frac{\Delta\tilde{\eta}_t}{GDP_{1990}} = -0.24\Delta X_{3t} + 3.00\Delta X_{4t} + 1.49\Delta X_{6t} + 1.01\Delta X_{7t} \quad (2)$$

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:

$$\frac{\tilde{\eta}_t}{GDP_{1990}} \times \frac{\eta^*_{1990}}{GDP_{1990}} \times \frac{GDP_{1990}}{\tilde{\eta}_{1990}} \times \frac{GDP_{1990}}{GDP_t} = \frac{\hat{\eta}_t}{GDP_t} \quad (3)$$

I. $\frac{\tilde{\eta}_t}{GDP_{1990}}$ is the index of shadow economy calculated by eq.(2);

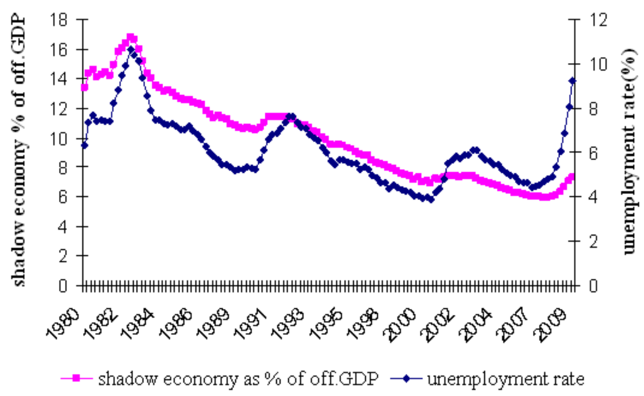
II. $\frac{\eta^*_{1990}}{GDP_{1990}} = 10.6\%$ is the exogenous estimate of shadow economy;

III. $\frac{\tilde{\eta}_{1990}}{GDP_{1990}}$ is the value of index estimated by eq.(2);

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. $\frac{\hat{\eta}_t}{GDP_t}$ is the estimated shadow economy as a percentage of official GDP.

Fig. 1. Shadow economy vs. Unemployment rate



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 ([10], [18], [19]).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. A nonparametric analysis of the relationship between unemployment rate and the size of the shadow economy using spline models

The paper aim to investigate the relationship between shadow economy estimated using the MIMIC model and unemployment rate using a nonparametric analysis based on spline models. The unemployment rate is expressed in %, taken from U.S. Bureau of Statistics, Labour Force Statistics from Current Population Survey.

Instead of assuming that we know the functional form for a regression model, a better alternative is to estimate the functional form from the data, replacing global estimates with local estimates. In the terms of local estimation, the statistical dependency between two variables is described not with a single parameter such as a mean or a slope coefficient, but with a series of local estimates.

Like local polynomial regression (LPR), spline smoothers are another nonparametric technique used with scatterplots. In any spline model, it must be selected the number of knots and the knot placement [16].

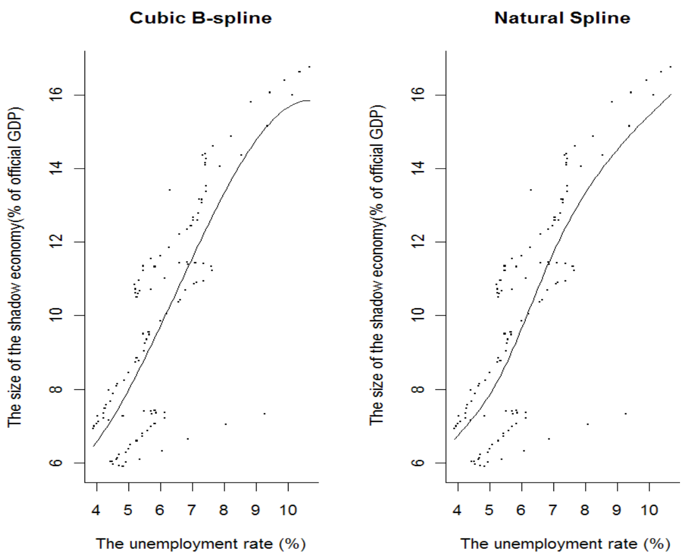
Standard practice is to place knots at evenly spaced intervals in the data. But the question of how to select the number of knots remains and has an important effect on the Spline fit. One method is to use a visual trial. Four knots is the standard starting point. If the fit appears rough, knots are added. If the fit appears overly nonlinear, knots are subtracted. The second method is to use Akaike Information Criterion to select the number of knots. The optimal number of knots is returned by the lowest AIC value.

Thus far, LPR estimates have revealed a linear dependency between the size of the shadow economy and the unemployment rate. It will we interesting to investigate the nature of the relationship between the two variables, using both cubic B-splines and natural cubic B-splines to estimate the nonparametric fit.

For the both spline models, it has been used 4 knots it we will evaluate whether this is the optimal number of knots, using Akaike Criterion.

Analyzing the graphics of both functions, there is a little difference between cubic B-splines and the natural cubic B-splines.

Fig.2. Cubic B-spline and natural spline fit to size of the shadow economy (% of off.GDP)



ISBN: 978-960-474-287-5

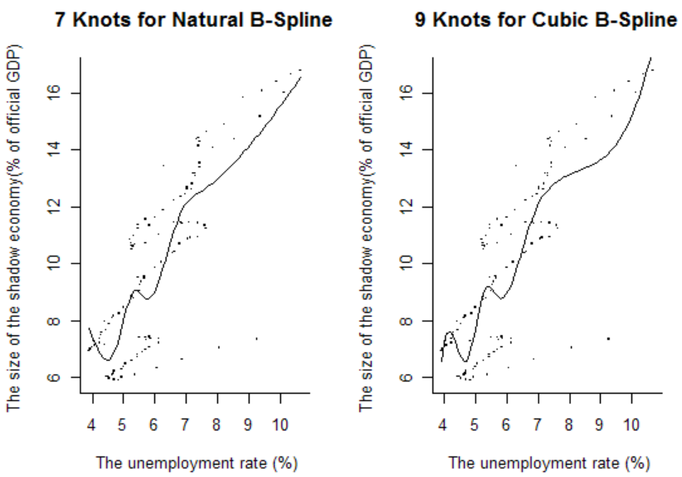
In order to select the number of knots, we use for both models the Akaike Information Criterion (AIC); the optimal number of knots is returned by the lowest AIC value. For the both spline models, it has been estimated several models with 2-9 knots.

Table 3. AIC values for differing number of knots

	Natural Spline	Cubic Spline
2 knots	477.0573	478.7849
3 knots	478.1010	478.7849
4 knots	477.0685	478.7849
5 knots	479.0302	478.7093
6 knots	480.0504	479.2769
7 knots	***475.1998***	477.6035
8 knots	477.5902	479.2186
9 knots	477.2197	***475.646***

Analysing the values of Akaike Information Criterion for several knots we observe that the optimal number of knots for the Natural Spline taking into account the lowest AIC value is 7 knots, while for the cubic Spline the optimal number of knots is 9.

Fig.3. The choice of the optimal number of knots for the both models



In order to estimate the statistical relationship between two variables, both splines and local polynomial regression can provide such an estimate with few assumptions about functional form. A common criticism of the both methods is that it is easy to have a surfeit of local parameters, which produces overly nonlinear estimates that overfit data [16].

Penalized splines are a nonparametric regression technique that minimizes the possibility of overfitting. Smoothing splines operate with penalized estimation, placing a penalty on the number of local parameters used to estimate the nonparametric fit.

Like linear regression models, the spline estimate $\hat{f}$ minimises the sum of squares between y and the nonparametric estimate, $f(x_i)$:

$$SS(f) = \sum_{i=1}^{n}[y - f(x)]^2 \qquad (4)$$

The main problem is that the estimate of f that minimises (4) use too many parameters. The penalised estimation solution is to attach a penalty for the number of parameters used to estimate f: $\lambda\int_{x_1}^{x_n}[f''(x)]^2dx$, named roughness penalty that have two components: λ, the smoothing parameter and the second, the integrated squared second derivative of $f(x)$.

Further, the spline estimate become:

$$SS(f,\lambda) = \sum_{i=1}^{n}[y - f(x)]^2 + \lambda\int_{x_1}^{x_n}[f''(x)]^2dx \qquad (5)$$

While small values of λ will interpolate the data and large values returns a least squares fit, intermediate values does not offer an interpretable effect on the amount of the smoothing applied to the data. It is proposed a transformation of λ into an approximation of the degrees of freedom; by selecting the degrees of freedom, it is chosen the number of effective local parameters used in the spline estimate.

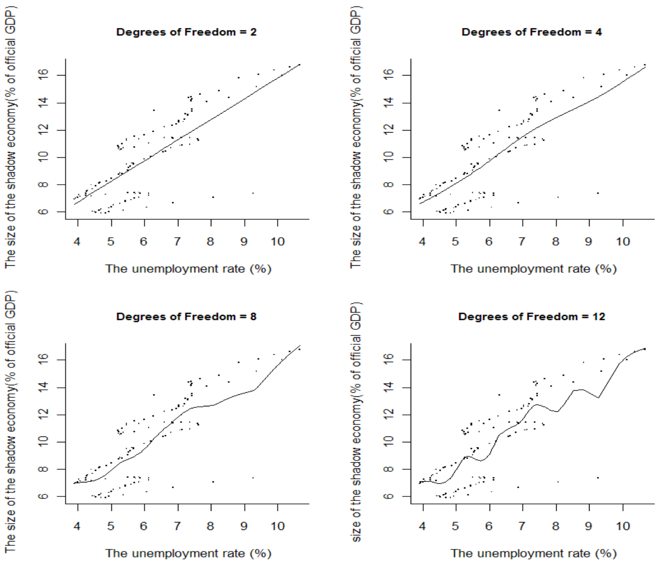
The penalized splines named also “smoothing splines” differ from the standard splines by the fact that the number of knots have little influence over how smooth the fit is since the value of λ controls now the quality of the fit.

In order to see how different degrees of freedom (2, 4, 8 and 12) affect the fit, it has been estimate the relationship between the size of the shadow economy and the unemployment rate using smoothing splines.

The results reveal that the fit with 2 degrees of freedom is identical to a linear regression; for the model with 4 and 8 degrees of freedom we have the same pattern of linearity found with other spline fits. For the fit with 12 degrees of freedom, we have considerable variability, caused by too many parameters and we can conclude that the data are over fitted.

ISBN: 978-960-474-287-5

Fig.4. Smoothing spline fit to shadow economy data



In order to test hypothesis about the nature of the relationship between the size of the shadow economy and the unemployment rate, we compare the smoothing spline model to a model with only a constant to test whether the effect of the unemployment rate is significantly different from zero.

If RSS_1 and RSS_2 are the residual sum of squares from a restricted model and the spline model respectively, the

$$F-test=\frac{(RSS_1-RSS_2)/(df_{res2}-df_{res1})}{RSS_2/(n-df_{res2})}\approx F_{df_{res2}-df_{res1},n-df_{res2}} \quad (6)$$

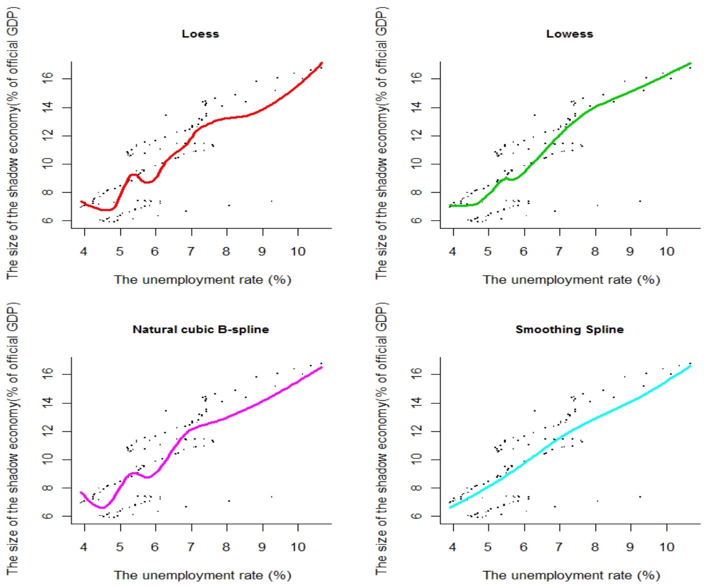
Applying the F-test, we find that the relationship between the two variables is highly significant as the test statistic is 72.718 on 3 and 115 degrees of freedom (p= 2.2e-16). We also test the spline model against a global linear fit, and the value of F-test of 1.9597 on 2 and 114 degrees of freedom is not statistically significant(p=0.1476). The results of the test indicate that the relationship between the size of the shadow economy and the unemployment rate is sufficiently linear and the global linear fit is adequate.

Finally, we provide a comparison of the nonparametric regression models (fig.5).

The first two nonparametric models are the loess and lowess smoothers. In Dobre, Alexandru [9] we have estimated the both models identifying the optimal value of span at the value of 0.4. Between the two LPR smoothers, the lowess estimate provides a better fit of the data.

In the lower left panel is the natural cubic B-spline with 7 knots, chosen by the AIC values. This model displays noticeable undersmoothing of the estimate. Finally, we estimate a smoothing spline using 4 degrees of freedom selected through visual trial. The smoothing spline closely matches the linearity between the size of the shadow economy and the unemployment rate.

Fig. 5. Comparison of smoother fits



4. Conclusions

The main goal of the paper is to investigate the nature of the relationship between unemployment rate and the size of the shadow economy of the USA data using spline models.The shadow economy is estimated as percentage of official GDP, using MIMIC model. The results show that the size of the shadow economy varies from thirteen to seventeen percent between 1980 and 1983 and then decreases steadily up to 7 percent of official GDP in 2009.

We investigate the nature of the relationship between the two variables, using cubic B-splines and natural cubic B-splines to estimate the nonparametric fit. The graphics of both spline models reveals a little difference between the two functions.

Using an F-test, we compare the smoothing spline model to a model with only a constant, and we conclude that unemployment rate has a statistically significant effect on the size of the U.S.A. shadow economy.

We also test the spline model against a global linear fit and the results indicate that the relationship between the size of the shadow economy and the unemployment rate is sufficiently linear and the global linear fit is adequate.

Finally, we have compared the local polynomial regression models (loess and lowess) estimated in

(Dobre, Alexandru, 2010) with spline models (natural cubic B-spline and smoothing spline). From the four types of models that we have applied, the smoothing spline model closely matches the linearity between the size of the shadow economy and the unemployment rate.

References

[1] Alexandru A., Dobre, I., Ghinararu, C. ”The relationship between shadow economy and unemployment rate: a SVAR approach” Proceedings of 5th WSEAS International Conference on Economy and Management Transformation, Timisoara, Romania, october 23-26, 2010, pg.242-249, ISSN: 1792-5983.
[2] Bertola, G., Garibaldi, P., *The Structure and History of Italian Unemployment*, *CESifo Working Papers*, n.907, 2003.
[3] Boeri, T., Garibaldi, P., *Shadow Activity and Unemployment in a Depressed Labor Market*, *CEPR Discussion papers,* n.3433, 2002.
[4] Dell'Anno, R., *Estimating the shadow economy in Italy: A structural equation approach*, Working Paper 2003, Department of Economics, University of Aarhus.
[5] Dell'Anno, R., Gomez, M., Alañón Pardo, A., *Shadow economy in three different Mediterranean countries: France, Spain and Greece. A MIMIC approach*, Empirical Economics 33/2005, pp. 51-84.
[6] Dell'Anno, R., Schneider, F., *The Shadow Economy of Italy and other OECD Countries: What do we know?*, Mimeo, 2004.
[7] Dobre I., Alexandru, A., *The impact of unemployment rate on the dimension of shadow economy in Spain: a Structural Equation Approach*, European Research Studies Journal, vol. XIII, no. 4/2009, pg.179-197, ISSN: 1108-2976.
[8] Dobre, I., Alexandru, A., *Estimating the size of the shadow economy in Japan: A structural model with latent variables*, Economic Computation and Economic Cybernetics Studies and Research, vol.43 no.1/2009, pg.67-82, ISSN 0424 – 267 X.
[9] Dobre, I., Alexandru, A, *A nonparametric analysis of the relationship between unemployment rate and shadow economy using local polynomial regression models*, Economic Computation and Economic Cybernetics Studies and Research, vol.44, no.1/2010, pg.21-44, ISSN 0424 – 267 X.
[10] Enste, D.H., *Shadow Economy and Institutional Change in Transition Countries* in Boyan Belev (eds.), *The Informal Economy in the EU Assessment Countries: Size, Scope, Trends and Challenges of the Process of EU-enlargement*, Center for Study of Democracy, 2003, Sofia, 81-114.
[11] Giles, D.E.A., *Measuring the hidden economy: Implications for econometric modeling*, The Economic Journal, vol.109, no. 456/1998 pp.370-380.
[12] Giles, D.E.A., Modeling *the hidden economy in the tax-gap in New Zealand*, Empirical Economics, vol.24, no.4/1999, pp.621-640.
[13] Johnson, S., Kaufmann, D., Zoido-Lobatón, P., *Regulatory discretion and the unofficial Economy*, The American Economic Review, vol.88, no.2/1998, pp. 387-392.
[14] Jöreskog K., Goldberger, A.S., *Estimation of a model with multiple indicators and multiple causes of a single latent variabl*, *Journal of the American Statistical Association*, 70/1975, pp.631-639.
[15] Jöreskog, K., Sörbom, D., *LISREL 8 User's Reference Guide* (Scientific Software International, Chicago), 1993.
[16] Keele, L. *Semiparametric Regression for the Social Sciences*. Wiley and Ltd, 2008.
[17] Lackó, M., *Hidden economy an unknown quantitiy? Comparative analyses of hidden economies in transition countries in 1989-95"*,Working paper 9905, Department of Economics, University of Linz, 1999.
[18] Schneider, F., Enste, D.H., *Shadow economies: size, causes and consequences*, Journal of Economic Literature 38/2000, pp. 77-114.
[19] Schneider, F., *Shadow Economies and Corruption all over the world: New estimates for 145 Countries,* Economics, 2009, pp. 1-47.
*** www.bea.gov , U.S. Economic Accounts
*** www.bls.gov , U.S. Department of Labour Statistics
*** Eviews 6.0 software
*** Lisrel 8.8 package
*** R 2.9.1 software