I Z A

IZA DP No. 3306

MIMIC Models, Cointegration and Error Correction: An Application to the French Shadow Economy

Andreas Buehn
Friedrich Schneider

January 2008

MIMIC Models, Cointegration and Error Correction: An Application to the French Shadow Economy

Andreas Buehn
Dresden University of Technology

Friedrich Schneider
*Johannes Kepler University of Linz
and IZA*

ABSTRACT

MIMIC Models, Cointegration and Error Correction: An Application to the French Shadow Economy

The analysis of economic loss attributed to the shadow economy has attracted much attention in recent years by both academics and policy makers. Often, multiple indicators multiple causes (MIMIC) models are applied to time series data estimating the size and development of the shadow economy for a particular country. This type of model derives information about the relationship between cause and indicator variables and a latent variable, here the shadow economy, from covariance structures. As most macroeconomic variables do not satisfy stationarity, long run information is lost when employing first differences. Arguably, this shortcoming is rooted in the lack of an appropriate MIMIC model which considers cointegration among variables. This paper develops a MIMIC model which estimates the cointegration equilibrium relationship and the error correction short run dynamics, thereby retaining information for the long run. Using France as our example, we demonstrate that this approach allows researchers to obtain more accurate estimates about the size and development of the shadow economy.

JEL Classification: O17, O5, D78, H2, H11, H26

Keywords: shadow economy, tax burden, regulation, unemployment, cointegration, error correction models, MIMIC models

Corresponding author:

Friedrich Schneider
Department of Economics
Johannes Kepler University of Linz
A-4040 Linz-Auhof
Austria
E-mail: friedrich.schneider@jku.at

1. Introduction

This paper presents a first attempt to econometrically improve the multiple indicators multiple causes (MIMIC) model in order to gather more precise information about the size and development of the shadow economy over time. Its main contribution is the first ever analysis of the cointegration of the shadow economy's causes and indictors and the consideration of their long and short run relationships.

The shadow economy is still a controversial issue. Every day, many people around the world are engaged in black market activities. These people usually go underground in order to avoid taxation and save money as products and services are considerably cheaper. Policymakers are concerned about rising shadow economic activities because they weaken tax and social security bases. As a result, they may make ineffective policy decisions based on erroneous official indicators, which in turn give rise to crowding-out effects of official economic activity (Schneider and Enste, 2000). Economists focus on the shadow economy to improve economic theory. They also try to obtain accurate measures of the shadow economy for more effective formulation of economic policies. Specifically, the size and development of the shadow economy is a fundamental requirement. Our paper presents an improved estimation procedure that provides more precise figures than previous attempts to quantify the shadow economy.

Theoretical and empirical research in the field of shadow economics has gained much attention[1]. Although some progress has been made, it is still a difficult task because individuals engaged in such activities do not wish to be identified. Consequently, the estimation of the shadow economy becomes a scientific passion to know the unknown. Today, a variety of techniques has been employed to measure the size and development of the shadow economy. This includes direct approaches such as surveys and discrepancy methods as well as indirect methodologies like the transaction or the currency demand approach.[2] Recently, the MIMIC model was derived as a special type of structural equation model with latent variables.[3] It analyzes the covariance structures between observable cause and indicator variables to derive information about the relationship between them and an unobservable

---

[1] See for example Schneider and Enste (2000 and 2002).
[2] See Schneider and Enste (2000) for a reliable survey about diverse approaches.
[3] The MIMIC model approach traces back to Weck (1983), Frey and Weck (1983), and Frey and Weck-Hannemann (1984) and was enhanced by Aigner *et al*. (1988).

latent variable. Because it is generally agreed that the shadow economy can be treated as a latent variable, the MIMIC model rightly supplements existing direct and indirect approaches. In contrast to the latter, its major advantage is that it differentiates between causes and indicators and, most notably, when estimating the shadow economy considers its various causes.

Often, MIMIC models are applied to time series data to derive estimates of the size and development of the shadow economy over time. As most macroeconomic variables do not satisfy the underlying assumption of stationarity, the problem of spurious regressions may arise. Researchers usually overcome this problem by transforming the time series into stationary ones, employing a difference operator. Alternatively, one could estimate an error correction model (ECM) if the variables were cointegrated and a stationary long run relationship existed between them. This approach has become popular in applied economics in recent years. The former approach, however, is still used in MIMIC model investigations of the shadow economy. The latter MIMIC model in first differences is referred to as the DYMIMIC model (Aigner *et al.*, 1988). In the DYMIMIC model, on the other hand, the data's long run information is lost if the variables are used in their first differences, albeit they are cointegrated. Retaining this information may help to improve estimates of the shadow economy. The authors are not aware of any study that tests for cointegration in the variables and makes use of their long run equilibrium relationships in a MIMIC model estimation.

The paper is organized as follows. Section 2 provides a short summary of the basics of cointegration and the error correction mechanism. The MIMIC model is presented in Section 3. Section 4 widens the traditional MIMIC model, allowing for cointegration of the variables, and presents the covariance matrices for the new model's long run and short run equations. In Section 5 we employ our model to France and present the results. Section 6 concludes.

2. Cointegration and Error Correction Models

Empirical investigations in macroeconomics using time series data mostly involve variables that do not fulfill the assumed statistical properties. Specifically, they are not stationary and integrated to an order d different from zero $(I(d), d > 0)$. In the past, such variables were differenced to remove random walk and/or trend components and then analyzed using the Box and Jenkins method. The drawback of this approach is that valuable information about

the variables' potential long run relationship is lost. Granger and Weiss (1983) later showed that two variables, x_t and y_t, each I(1), may have a linear combination, $u_t = y_t - \beta x_t$, that is, I(0). We would normally expect the standard linear regression model's error term, u_t, to be I(1). But if u_t is still a white noise series, x_t and y_t are said to be cointegrated with the cointegration vector $[1, -\beta]$. Generally, the cointegrated relation between variables is interpreted as their long run equilibrium. This concept of cointegration immediately enjoyed great popularity because it finally introduced the equilibrium concept of economics to econometrics.

For any two cointegrated I(1) variables x_t and y_t with cointegration vector $[1, -\beta]$, we know that their first differences Δx_t and Δy_t as well as their long-run relation $y_t - \beta x_t$ are I(0). We can thus formulate the following equation:

$$\Delta y_t = \gamma \Delta x_t + \lambda u_{t-1} + w_t, \tag{1}$$

where w_t is a white noise error term and all expressions are I(0). The residual u_{t-1} is the one-period lagged equilibrium error of the cointegrated long run equation and is used as an error correction term in the dynamic, first difference regression of equation (1). Engle and Granger (1987) proved that this error correction model (ECM) is the data generating process of any two cointegrated variables x_t and y_t and that ECMs generate cointegrated variables.[4] Unlike the examination of the long run equilibrium, the ECM studies the relationship between the deviations of x_t and y_t from their respective long run trends. The expansion of equation (1) to the multivariate case where $\mathbf{y_t}$ is a vector of variables that are individually I(1), and $\mathbf{x_t}$ is a vector of variables containing a constant – variables that are assumed to be I(0) and where a time trend might also apply (Greene, 2007).

The concepts of cointegration and error correction enable researchers to study both the long run relationship between variables and the deviations from their respective long run trends and gather a better understanding of the economy's major features. Before showing how to utilize the idea of cointegration in an analysis of structural equation relations, we briefly introduce the standard MIMIC model.

---

[4] This class of models traces back to Sargan (1964), who introduced a model that retained long run information in a non-integrated specification for the first time.

3. The MIMIC Model

The MIMIC model explains the relationship between observable variables and an unobservable variable by minimizing the distance between the sample covariance matrix and the covariance matrix predicted by the model. The observable variables are divided into causes of the latent variable and its indicators. Formally, the MIMIC model consists of two parts: the structural equation model and the measurement model. The structural equation model is given by:

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

where $\mathbf{x}'_t = (x_{1t}, x_{2t}, \ldots, x_{qt})$ is a $(1 \times q)$ vector of time series variables as indicated by the subscript t. Each time series $x_{it}, i = 1, \ldots, q$ is a potential cause of the latent variable η_t. $\gamma' = (\gamma_1, \gamma_2, \ldots, \gamma_q)$, a $(1 \times q)$ vector of coefficients in the structural model describing the "causal" relationships between the latent variable and its causes. Since the structural equation model only partially explains the latent variable η_t, the error term ς_t represents the unexplained component. The MIMIC model assumes that the variables are measured as deviations from their means and that the error term does not correlate to the causes, i.e. $\mathbf{E}(\eta_t) = \mathbf{E}(\mathbf{x_t}) = \mathbf{E}(\varsigma_t) = 0$ and $\mathbf{E}(\mathbf{x_t}\varsigma'_t) = \mathbf{E}(\varsigma_t \mathbf{x}'_t) = 0$. The variance of ς_t is abbreviated by ψ and $\mathbf{\Phi}$ is the $(q \times q)$ covariance matrix of the causes x_t.

The measurement model represents the link between the latent variable and its indicators, i.e. the latent unobservable variable is expressed in terms of observable variables. It is specified by:

$$\mathbf{y_t} = \lambda \eta_t + \varepsilon_t, \tag{3}$$

where $\mathbf{y}'_t = (y_{1t}, y_{2t}, \ldots, y_{pt})$ is a $(1 \times p)$ vector of individual time series variables $y_{jt}, j = 1, \ldots, p$. $\mathbf{\varepsilon_t} = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{pt})$ is a $(p \times 1)$ vector of disturbances where every $\varepsilon_{jt}, j = 1, \ldots, p$ is a white noise error term. Their $(p \times p)$ covariance matrix is given by $\mathbf{\Theta_\varepsilon}$. The single $\lambda_j, j = 1, \ldots, p$ in the $(p \times 1)$ vector of regression coefficients λ, represents the magnitude of the expected change of the respective indicator for a unit change in the latent variable. Like the MIMIC model's causes, the indicators are directly measurable and expressed as deviations from their means, that is, $\mathbf{E}(\mathbf{y_t}) = \mathbf{E}(\mathbf{\varepsilon_t}) = 0$. Moreover, it is assumed that the error terms in the measurement model do not correlate either to the causes $\mathbf{x_t}$ or to

the latent variable η_t, hence, $\mathbf{E}(\mathbf{x_t \varepsilon_t'}) = \mathbf{E}(\mathbf{\varepsilon_t x_t'}) = 0$ and $\mathbf{E}(\eta_t \mathbf{\varepsilon_t'}) = \mathbf{E}(\mathbf{\varepsilon_t} \eta_t') = 0$. A final assumption is that the $\mathbf{\varepsilon_t}$ s do not correlate to ζ_t, i.e. $\mathbf{E}(\mathbf{\varepsilon_t} \varsigma_t') = \mathbf{E}(\varsigma_t \mathbf{\varepsilon_t'}) = 0$. Figure 3.1 shows the general structure of the MIMIC model.

Causes   **Indicators**
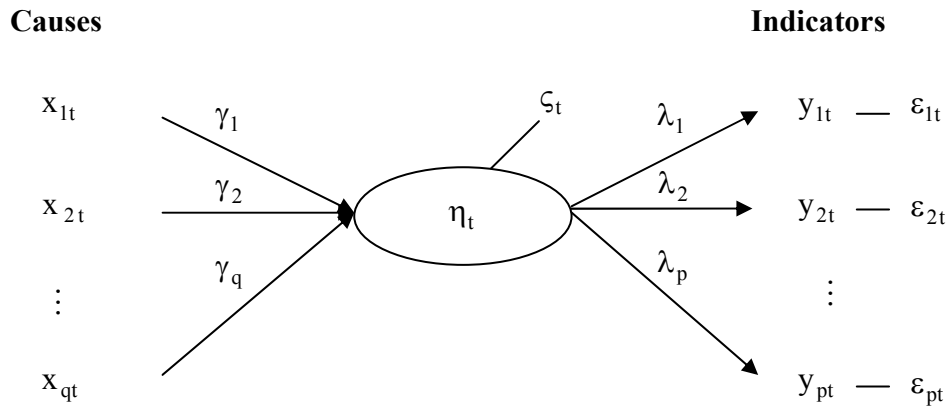
Fig. 3.1 *General Structure of a MIMIC Model*

From equation (2) and (3) and making use of the definitions, we can derive the MIMIC model's covariance matrix $\mathbf{\Sigma}$ (see Appendix A). This matrix describes the relationship between the observed variables in terms of their covariances. Decomposing the matrix derives the structure between the observed variables and the unobservable, latent variable, here, the shadow economy. The model's covariance matrix is given by:

$$\mathbf{\Sigma} = \begin{pmatrix} \lambda(\gamma'\mathbf{\Phi}\gamma + \psi) + \mathbf{\Theta_\varepsilon} & \lambda\gamma'\mathbf{\Phi} \\ \mathbf{\Phi}\gamma\lambda' & \mathbf{\Phi} \end{pmatrix}, \tag{4}$$

where $\mathbf{\Sigma}$ is a function of the parameters λ, γ and the covariances contained in $\mathbf{\Phi}$, $\mathbf{\Theta_\varepsilon}$, and ψ. Since the latent variable is not observable, its size is unknown, and the parameters of the model must be estimated using the links between the observed variables' variances and covariances. Thus, the goal of the estimation procedure is to find values for the parameters and covariances that produce an estimate for $\mathbf{\Sigma}$ that is as close as possible to the sample covariance matrix for the observed causes and indicators, i.e. the $\mathbf{x_t}$ s and $\mathbf{y_t}$ s. Now we extend this model to include the concept of cointegration.

4. The EMIMIC Model

4.1. The MIMIC Model and Co-integration

The first step in deriving the error correction MIMIC (EMIMIC) model is to put equation (2) into equation (3). This yields:

$$\mathbf{y_t} = \mathbf{\Pi x_t} + \mathbf{z_t} \ , \tag{5}$$

where $\mathbf{\Pi} = \lambda \gamma'$ and $\mathbf{z_t} = \lambda \zeta_t + \mathbf{\varepsilon_t}$. The error term $\mathbf{z_t}$ in equation (5) is a $(p \times 1)$ vector of linear combinations of the white noise error terms ς_t and $\mathbf{\varepsilon_t}$ from the structural equation and the measurement model, i.e. $\mathbf{z_t} \sim (\mathbf{0}, \mathbf{\Omega})$. The covariance matrix $\mathbf{\Omega}$ is given as $\mathbf{Cov(z_t)} = \lambda \lambda' \psi + \mathbf{\Theta_\varepsilon}$. Equation (5) is comparable to a simultaneous regression model where the endogenous variables y_{jt}, $j = 1, \dots, p$ are the latent variable η's indicators and the exogenous variables x_{it}, $i = 1, \dots, q$ its causes. The application of cointegration theory to the MIMIC model is thus possible.

From Section 2 we already know that the presumption that every element z_j, $j = 1, \dots, p$ of $\mathbf{z}$ is a stationary white noise series is probably untrue if some of the causes and indicators x_{it}, $i = 1, \dots, q$ and y_{jt}, $j = 1, \dots, p$ are I(1) series. Thus, if $\mathbf{x_t}$ and $\mathbf{y_t}$ are vectors of I(1) variables, one would normally expect every linear combination $y_{jt} - \boldsymbol{\pi_{j.}} \mathbf{x_t}$ to be I(1), i.e. a trend over time. Here, $\boldsymbol{\pi_{j.}}$ is the jth $(1 \times q)$ row vector of matrix $\mathbf{\Pi}$ in equation (5); therefore, the jth long run equilibrium error z_{jt} will also be I(1), which implies inconsistency in equation (5) of the model. On the other hand, we know that if a linear combination $z_{jt} = y_{jt} - \boldsymbol{\pi_{j.}} \mathbf{x_t}$ exists, where z_{jt} is I(0), the variables are said to be cointegrated (Engle and Granger, 1987). The vector $[1, -\boldsymbol{\pi_{j.}}]$, where $\boldsymbol{\pi_{j.}}$ itself is a $(1 \times q)$ row vector, is a cointegration vector. Generally, since every linear combination z_{jt}, $j = 1, \dots, p$ consists of $q + 1$ variables, there could be more than one cointegration vector – in principle up to q linearly independent cointegration vectors (Greene, 2007). If p indicators are I(1), the number of linearly independent cointegration vectors is $(p \cdot q)$.

Of course, not every macroeconomic variable is I(1): there may also be some I(0) time series. We therefore generalize the assumptions of the previous paragraph and denote these causes as a vector $\mathbf{v}_t' = (v_{1t}, v_{2t}, \ldots, v_{rt})$. Equation (5) then becomes:

$$\mathbf{y}_t = \mathbf{\Pi}\mathbf{x}_t + \mathbf{T}\mathbf{v}_t + \mathbf{z}_t, \qquad\qquad (6)$$

where $\mathbf{T} = \lambda\boldsymbol{\beta}'$ and $\boldsymbol{\tau}' = (\tau_1, \tau_2, \ldots, \tau_r)$ is the $(1 \times r)$ vector of coefficients of the I(0) variables in the structural relationship. Given that r of the causes are I(0), the dimensions of $\mathbf{\Pi} = \lambda\boldsymbol{\gamma}'$ and $\mathbf{x}_t$ are $[1 \times (q-r)]$ and $[(q-r) \times 1]$, respectively, and the dimensions of $\mathbf{T}$ and $\mathbf{v}_t$ are $(p \times r)$ and $(r \times 1)$, respectively. Consequently, if r of the causes are I(0), the maximum number of linearly independent cointegration vectors for every disturbance term z_{jt}, $j = 1, \ldots, p$ in equation (6) is $(q-r)$. Moreover, if $s \leq p$ of the p indicator variables are also individually I(0), the maximum number of linearly independent cointegration vectors decreases to $(q-r) - s$.

4.2. Error Correction Representation of the EMIMIC Model

As equation (1) shows, every cointegration relationship has an error correction mechanism where the long run relationship leads to equilibrium and the short run relationship contains a dynamic mechanism (Engle and Granger, 1987). Thus, equation (6) can be written as:

$$\Delta\mathbf{y}_t = \mathbf{A}\Delta\mathbf{x}_t + \mathbf{T}\mathbf{v}_t + \mathbf{K}\mathbf{z}_{t-1} + \mathbf{w}_t. \qquad\qquad (7)$$

where $\Delta\mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$, $\Delta\mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_{t-1}$, $\mathbf{z}_{t-1} = \mathbf{y}_{t-1} - \mathbf{\Pi}\mathbf{x}_{t-1}$, and $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{K}$ are coefficient matrices in this dynamic, short run model specification. Furthermore, in this specification $\mathbf{A} = \lambda\boldsymbol{\alpha}'$ is the $[p \times (q-r)]$ coefficient matrix of the first differences of the I(1) causes, and $\mathbf{B} = \lambda\boldsymbol{\beta}'$ is the $(p \times r)$ coefficient matrix of the I(0) causes. The matrix $\mathbf{K} = \lambda\boldsymbol{\kappa}'$ is the $(p \times p)$ coefficient matrix for the long run disequilibrium's error correction term and $\mathbf{w}_t \sim (\mathbf{0}, \boldsymbol{\Omega})$ is a white noise disturbance.[5] As long as $\mathbf{x}_t$ and $\mathbf{y}_t$ are vectors of I(1) variables, their changes will be I(0). Thus, every term in equation (7) – given that $\mathbf{y}_t$ and $\mathbf{x}_t$ in equation (6) are cointegrated, leading to $\mathbf{z}_t$ that is also individually I(0) – is I(0). Together, equation (6) and (7) define the EMIMIC model. Since the fundamental idea of the MIMIC model is to

---

[5] Both separate equations of the MIMIC model in first differences are $\Delta\mathbf{y}_t = \lambda\Delta\eta_t + \boldsymbol{\varepsilon}_t$ and $\Delta\eta_t = \boldsymbol{\alpha}'\Delta\mathbf{x}_t + \boldsymbol{\beta}'\mathbf{v}_t + \boldsymbol{\kappa}'\mathbf{z}_{t-1} + \varsigma_t$. Putting the first equation into the second and using the definitions of $\mathbf{A}, \mathbf{B}, \mathbf{K},$ and $\mathbf{w}_t = \lambda\varsigma_t + \boldsymbol{\varepsilon}_t$ gives, at first, $\Delta\mathbf{y}_t = \lambda\boldsymbol{\alpha}'\Delta\mathbf{x}_t + \lambda\boldsymbol{\beta}'\mathbf{v}_t + \lambda\boldsymbol{\kappa}'\mathbf{z}_{t-1} + \mathbf{w}_t$, and, finally, equation (7).

minimize the distance between the sample covariance matrix and the covariance matrix predicted by the model, the next step is to derive the underlying structural equation system, i.e. the covariance matrices of equations (6) and (7) for the EMIMIC model.

To begin, we define (6) in terms of $\boldsymbol{\Sigma}$. Its general structure is:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{Var(y_t)} \\ \mathbf{Cov(x_t, y_t)} & \mathbf{Var(x_t)} \\ \mathbf{Cov(v_t, y_t)} & \mathbf{Cov(v_t, x_t)} & \mathbf{Var(v_t)} \end{pmatrix}.$$

Next, we formulate this covariance matrix in terms of the observed variables as a function of the model parameters using the assumption established in Section 3 (see Appendix B). From this, it follows that:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\lambda}\left(\boldsymbol{\gamma}'\boldsymbol{\Phi}_1\boldsymbol{\gamma} + 2\boldsymbol{\gamma}'\mathbf{N}\boldsymbol{\tau} + \boldsymbol{\tau}'\boldsymbol{\Phi}_2\boldsymbol{\tau}\right)\boldsymbol{\lambda}' + \boldsymbol{\Theta}_\varepsilon \\ \left(\boldsymbol{\Phi}_1\boldsymbol{\gamma} + \mathbf{N}\boldsymbol{\tau}\right)\boldsymbol{\lambda}' & \boldsymbol{\Phi}_1 \\ \left(\mathbf{N}'\boldsymbol{\gamma} + \boldsymbol{\Phi}_2\boldsymbol{\tau}\right)\boldsymbol{\lambda}' & \mathbf{N}' & \boldsymbol{\Phi}_2 \end{pmatrix}, \tag{8}$$

where $\mathbf{N} := \mathbf{E}(\mathbf{v}_t \mathbf{x}_t')$. $\boldsymbol{\Phi}_1$ is the covariance matrix for the I(1) causes and $\boldsymbol{\Phi}_2$ for the I(0) causes. For all other variables, the definitions introduced in the previous sections still hold. The covariance matrix of the short run error correction mechanism for equation (7) is:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{Var(\Delta y_t)} \\ \mathbf{Cov(\Delta x_t, \Delta y_t)} & \mathbf{Var(\Delta x_t)} \\ \mathbf{Cov(v_t, \Delta y_t)} & \mathbf{Cov(v_t, \Delta x_t)} & \mathbf{Var(v_t)} \\ \mathbf{Cov(z_{t-1}, \Delta y_t)} & \mathbf{Cov(z_{t-1}, \Delta x_t)} & \mathbf{Cov(z_{t-1}, v_t)} & \mathbf{Var(z_{t-1})} \end{pmatrix}.$$

Taking $\mathbf{z_t} \sim (\mathbf{0}, \boldsymbol{\Omega})$ into account, defining $\mathbf{M} := \mathbf{E}(\mathbf{v}_t \Delta\mathbf{x}_t')$, and making use of the assumptions established, we can also derive the short run equation's covariance matrix in terms of the model's parameters and covariances (see Appendix C). Thus, the covariance structure of equation (7) is:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\lambda}\left(\boldsymbol{\alpha}'\boldsymbol{\Phi}_3\boldsymbol{\alpha} + 2\boldsymbol{\alpha}'\mathbf{M}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{\Phi}_2\boldsymbol{\beta} + \boldsymbol{\kappa}'\boldsymbol{\Omega}\boldsymbol{\kappa}\right) \\ \left(\boldsymbol{\Phi}_3\boldsymbol{\alpha} + \mathbf{M}'\boldsymbol{\beta}\right)\boldsymbol{\lambda}' & \boldsymbol{\Phi}_3 \\ \left(\mathbf{M}\boldsymbol{\alpha} + \boldsymbol{\Phi}_2\boldsymbol{\beta}\right)\boldsymbol{\lambda}' & \mathbf{M} & \boldsymbol{\Phi}_2 \\ \boldsymbol{\lambda}\left(\boldsymbol{\psi}\boldsymbol{\lambda}'\boldsymbol{\kappa} + \boldsymbol{\psi}\right)\boldsymbol{\lambda}' + \boldsymbol{\Theta}_\varepsilon & \mathbf{0} & \mathbf{0} & \boldsymbol{\Omega} \end{pmatrix}. \tag{9}$$

A closer examination of equations (8) and (9) and a comparison with the MIMIC model's covariance matrix in equation (4) shows the effect of cointegration. In contrast to the latter, the EMIMIC model's covariance matrix in (9) is adjusted by the long run equilibrium error term's covariance matrix $\boldsymbol{\Omega}$ and the error correction term's parameter vector $\boldsymbol{\kappa}$. The further

modifications are needed to separate the I(0) and I(1) causes forming sub-matrices $(\mathbf{\Phi_3\alpha + M'\beta})\mathbf{\lambda'}$, $(\mathbf{M\alpha + \Phi_2\beta})\mathbf{\lambda'}$, $\mathbf{\Phi_2}$, $\mathbf{\Phi_3}$, and $\mathbf{M}$. However, because $\mathbf{\Sigma}$ is still a function of the model parameters $\mathbf{\alpha}$, $\mathbf{\beta}$, $\mathbf{\kappa}$, and $\mathbf{\lambda}$ and of covariances, the estimation procedure ensures that an estimate for the EMIMIC model's covariance matrix can be adapted. This model is now applied to estimate the size and development of the French shadow economy.

5. Applying the EMIMIC Model

5.1 The Concrete EMIMIC Model

Most analyses of the shadow economy come to the conclusion that tax and social security burdens and the intensity of regulation are the two main causes affecting the size and development of the shadow economy.[6] Taxes affect labor-leisure choices and stimulate labor supply in the shadow economy since the greater the difference between the total cost of labor in the official economy and the after-tax earnings from work, the greater the incentive to avoid this difference and to work in the shadow economy. An increase in the intensity of regulations, such as trade barriers and labor restrictions for foreigners, reduces the freedom (of choice) for individuals engaged in the official economy and leads to a substantial increase in labor costs in the official economy. Since most of these costs can be shifted onto employees, there is further incentive to work in the shadow economy – where they can be avoided. Statistically significant empirical evidence of the influence of taxation and the intensity of regulation on the shadow economy is provided in studies by Schneider (1994, 2005) and Johnson *et al*. (1998a, 1998b).

Unemployment and the hours worked per employee in the official economy also affect the shadow economy. While it is clear that a reduction in working hours in the official economy increases hours spent working in the shadow economy, unemployment's effect on the shadow economy is ambiguous. Whether the unemployment variable exhibits a positive or negative relationship depends on income and substitution effect. Income losses due to unemployment reduce demand in the shadow as well as the official economy. A substitution takes place as

---

[6] See Thomas (1992), Lippert and Walker (1997), Schneider (1994, 1997, 2003, 2005), Johnson et al. (1998a,b), Tanzi (1999), Giles (1999), Mummert and Schneider (2001), Giles and Tedds (2002), Giles et al. (2002), and Dell'Anno and Schneider (2003).

unemployed workers turn to the shadow economy where cheaper goods make it easier to countervail utility losses. This may stimulate additional demand there. If the income effect exceeds the substitution effect, a negative relationship develops. Likewise, if the substitution effect exceeds the income effect, the relationship is positive.

To mirror activities in the shadow economy, we use the monetary aggregate M1 and a GDP volume index. These variables are particularly suitable for this purpose as a result of the following considerations. Transactions in the shadow economy are typically carried out using cash or money that is drawn from a current account at a moment's notice. We therefore expect a positive relationship between the shadow economy and M1. The lower the officially measured GDP, the fewer possibilities people have to earn money in the official economy, and the likelier they are to be driven into the shadow economy. In the short run, we expect this negative relationship to exist. In the long run, however, the official and the shadow economy are complements rather than substitutes, and the variables will thus exhibit a positive relationship. One possible explanation is that when the official economy grows, the shadow economy grows as well since favorable economic conditions do not discern between the official and unofficial economy. The demand for maintenance and other services in the shadow economy in particular increase in the long run as a result of higher consumption (e.g. cars) in the official economy. Based on these theoretical considerations, we employ tax and social security contribution burdens, the intensity of regulation (measured by the ratio of government employment to the total labor force), the unemployment rate, and an index of hours worked per employee as causal variables for the shadow economy in the long run equilibrium estimation. A GDP volume index and the monetary aggregate M1 are used as indicators. Figure 5.1 illustrates these relationships. The small squares attached to the arrows indicate the expected signs in the empirical analysis.
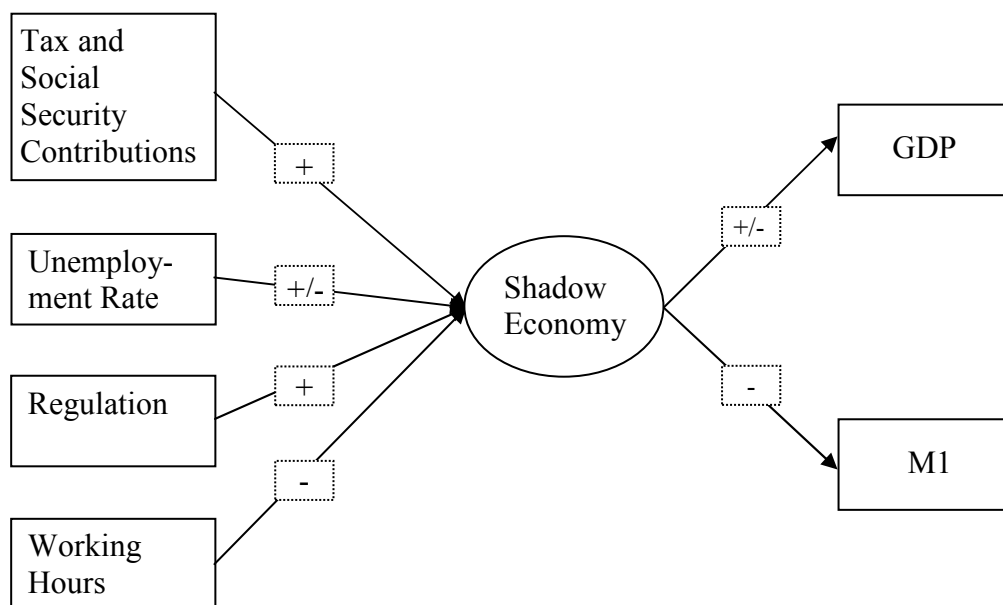
Fig. 5.1 *Hypothesized Relationships in the MIMIC Model*

5.2 Data and the Analysis of Stationarity and Co-integration

Our data cover each quarter between 1981 and 2006: the number of observations is 104. Data sources and in-depth definitions of the variables are summarized in Table D.1 in the Appendix. We begin our empirical analysis by pre-testing the data. In the first step, each series is individually examined under the null hypothesis of a unit root against the alternative of stationarity. As shown in Table 5.1 we find that all variables are I(1) using conventional unit root tests.

Table 5.1: *Analysis of Stationarity*

Variable	Test Equation	Level			Test Equation	First Difference		
Causes		ADF	PP	KPSS		ADF	PP	KPSS
Tax	C & T	0.2707	0.4633	0.103[***]	C	0.0000	0.0000	0.052[***]
Unemp	C & T	0.5072	0.6582	0.196[*]	C	0.0106	0.0006	0.304[***]
Reg	C & T	0.2000	0.3190	0.276	C	0.0023	0.0012	0.539[*]
Work	C & T	0.4872	0.2755	0.145[**]	C	0.0286	0.1712	0.138[***]
Indicators		ADF	PP	KPSS		ADF	PP	KPSS
GDP	C & T	0.1696	0.5827	0.118[*]	C	0.0009	0.0000	0.109[***]
M1	C & T	0.8536	0.0882	0.188	C	0.0208	0.0000	0.179[***]

[*] Stationarity at 1%. [**] Stationarity at 5%. [***] Stationarity at 10%.
Note: For the Augmented Dickey-Fuller (ADF) test and the Phillips-Perron (PP) test, the MacKinnon (1996) one-sided *p* values are given whereas test statistics are reported for the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. Its critical values are taken from Kwiatkowski *et al.* (1992). For a test equation with constant (C) the critical values are: 0.347 (10% level), 0.463 (5% level), and 0.739 (1% level) whereas for a test equation with constant and trend (C & T) the critical values are: 0.119 (10% level), 0.146 (5% level), and 0.216 (1% level). For the order of the autoregressive correction for the ADF test, we use the Akaike Information Criterion (AIC). For the PP and KPSS tests, we use the Bartlett kernel estimator and the Newey-West (1994) data-based automatic bandwidth parameter method.

Next, we use the Engle and Granger two-step approach to see if all four causes are cointegrated with each indicator variable and therefore exhibit a valid error correction representation (Engle and Granger, 1987). To do this, we first estimate least square regressions with variables in levels, where the particular indicator is the dependent variable and the causes are the independent variables. Thus, the regression equations are:

$$GDP = \alpha_1 \cdot \text{Taxes} + \alpha_2 \cdot \text{Unemp} + \alpha_3 \cdot \text{Reg} + \alpha_4 \cdot \text{Work} + u_1 \text{ and}$$

$$M1 = \alpha_1 \cdot \text{Taxes} + \alpha_2 \cdot \text{Unemp} + \alpha_3 \cdot \text{Reg} + \alpha_4 \cdot \text{Work} + u_2 \text{ .}$$

Because all variables are deviations from their means, no constant is included in the regression equations. Next we analyze the assumed cointegration relationship's residuals u_1 and u_2 using the ADF test. If the causes are cointegrated with the indicators, we expect the ADF test to reject the null hypothesis of a unit root against the alternative for both error terms u_1 and u_2. As presented in Table 5.2, we can in fact reject the null hypothesis for both residuals at conventional significance levels. We therefore conclude that the causes are cointegrated with each indicator.

Table 5.2: *Analysis of Cointegration Between Causes and Indicators*

Indicators	Causes				*t*-statistic for Residual	Jarque Bera Probability
GDP	Tax (0.0006)	Unemp (0.0000)	Reg (0.0007)	Work (0.0000)	-4.0569	0.55
M1	Tax (0.0452)	Unemp (0.0119)	Reg (0.0039)	Work (0.0000)	-3.8725	0.62

Note: The critical values of the ADF test's *t*-statistic are taken from Engle and Yoo (1987). For a sample with 100 observations, they are: 4.61 (1% level), 4.02 (5% level), and 3.71 (10% level). The order of the autoregressive correction has been chosen using the AIC as suggested by Engle and Yoo (1987). Thus, the null hypothesis of a unit root is rejected at the 10% level for residual u_1 and at the 5% level for residual u_2. The *p*-values of the parameter estimators are given in parenthesis.

The confirmation of both cointegration relationships permits the estimation of a long run equilibrium MIMIC model for the size and development of the shadow economy according to equation (5). The next step is to estimate the short run MIMIC model of equation (7) employing first differences of all causes and indicators. The estimation also includes the long run error correction terms u_1 and u_2 from both cointegration relationships. Here, the number of observations is 103. Table 5.3 presents the long run equilibrium model's parameter estimates and primary test statistics as well as those for the short run model. Supplementary summary statistics indicating the overall fit of both MIMIC models are provided in Table D.2 in the Appendix.

Table 5.3: *MIMIC Models and Parameter Estimates*

	Long-run MIMIC Model	Short-run MIMIC Model
Causes		
Taxes	0.17^{***}	0.50^{***}
	(5.74)	(6.84)
Unemp	-0.20^{***}	0.26^{***}
	(-5.14)	(3.47)
Reg	0.22^{***}	0.21^{***}
	(5.39)	(2.86)
Work	-0.75^{***}	-0.03
	(-21.47)	(-0.40)
u_1	--	-0.04
		(-0.47)
u_2	--	-0.27^{***}
		(-3.50)
Indicators		
GDP	1.00	-1.00
M1	0.99^{***}	0.19
	(45.58)	$(1.43)^{*}$
Statistics		
Chi-square	2.54	17.17
Degrees of Freedom	13	26
P-value	0.9989	0.9073
Root Mean Squared Error of Approximation (RMSEA)	0.000	0.000

T-statistics are reported in parenthesis. * Significance at 10% level. ** Significance at 5 % level. *** Significance at 1% level.

The estimated coefficients of all variables in the long run equilibrium relationship are highly statistically significant at the 1% significance level and have the theoretically expected sign. In the short run, the MIMIC model estimates for the independent variable "Work" and the residual u_1 from the cointegration relationship between the causes and the GDP are not statistically significantly different from zero. As in the long run equilibrium relationship, all remaining variables – with the exception of M1, which is statistically significant only at the 10% level – are statistically significant at the 1% level. Overall, the various test statistics point to a close fit.

In order to estimate not only the relative size of the parameters but also their levels, it is necessary to fix a scale for the unobservable latent variable. A convenient way to determine the relative magnitude of the variables is to set the coefficient of one of the measurement model's indicator variables to non-zero.[7] Here, we fix the coefficient of the variable GDP for both the long run and the short run MIMIC estimation.

We now summarize our findings from the estimations. First, we can confirm our theoretical considerations. Second, tax and social security contribution burdens and the intensity of regulation are important causes of the size of the shadow economy. These findings confirm our hypothesis as well as previous empirical findings. As expected, the relation between hours worked per employee and the shadow economy is negative, i.e. decreasing working hours encourage people to engage in informal economic activities. With regard to unemployment, it turns out that the overall long run effect is negative. That is, the income effect exceeds the substitution effect. This finding is supported by the positive relation between the shadow economy and official GDP, suggesting that in the long run the two are complements rather than substitutes. As expected, however, the short run relationship is negative, i.e. people who face unemployment switch to the shadow economy thereby negatively affecting official GDP. The other causes also show the expected sign in the short run estimation. That is, high tax and social security contribution burdens and a high intensity of regulation force people into the shadow economy. Declining working hours create the required freedom for those activities.

Both the long run equilibrium relationship and the short run dynamic error correction representation of the MIMIC model represent our EMIMIC model. With it, we can now estimate the size and development of France's shadow economy. The first step uses the long run part of our model to calculate the ordinal index. This index is then transformed into a cardinal series using the average of the estimates from Dell' Anno et al. (2007) and Schneider (2005), which is 14.25% of official GDP in 1995. Next, the short run deviations from equilibrium are calculated. Finally, taking these into account, estimates for the French shadow economy are derived using Bajada and Schneider's (2005) calibration methodology. Figure 5.2 illustrates these results. The thick line is the long run equilibrium, and the thin line represents the final estimates for the French shadow economy taking short run dynamic fluctuations into account.
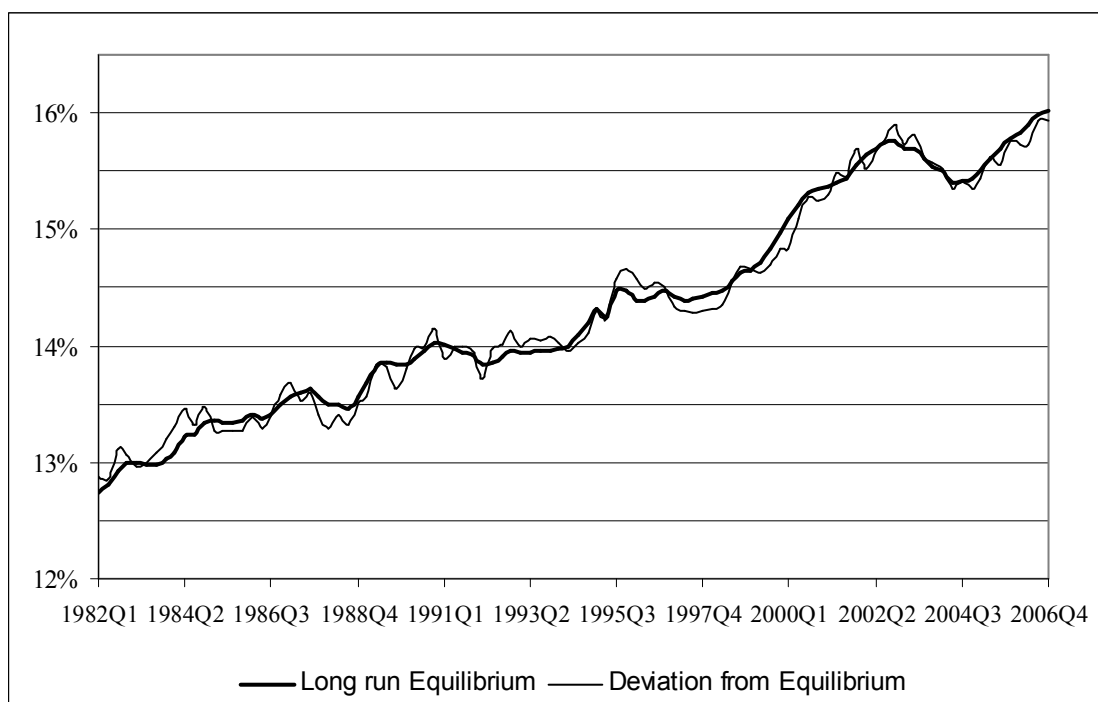
---

[7] Giles and Tedds (2002), 109.

Fig. 5.2 *Size of the French Shadow Economy in % of Official GDP*

Applying the EMIMIC model to the French shadow economy produces an estimate of 12.88% in Q1 1982, which increases to 15.93% in Q4 2006. All quarterly estimates for 1982 to 2006 are shown in Table D.3 in the Appendix. Table D.4 compares these estimates with those of the long run equilibrium relationship and illustrates this deviation for France's shadow economy. Thus, our findings correspond to other recent studies of the French shadow economy with 13.80% for 1990-93 (Schneider and Enste, 2000) and 15.30% for 2000 (Schneider, 2005). Our approach is nonetheless quite different from previous investigations because we use the long run equilibrium estimation for the initial calculation of the cardinal time series index. The calibration methodology is then only used to correct for deviations from equilibrium in the short run. Previous studies, on the other hand, derive the cardinal index of the shadow economy in a particular country from their DYMIMIC estimates using some type of calibration methodology. Our EMIMIC model thus estimates the size and development of the shadow economy more precisely.

6. Summary and Conclusions

In this paper we consider cointegration and error correction techniques within a latent variable approach. First, we demonstrate the possibility of controlling for long run equilibrium relationships in a MIMIC model by using the standard econometrics of cointegration and error correction mechanisms. Next, we present the cointegration equation and the error correction equation of a MIMIC model where indicators and causes are cointegrated. Finally, we derive the covariance matrices for the EMIMIC model and employ this model to estimate the size and development of the French shadow economy. We demonstrate that the EMIMIC model better quantifies the size of the shadow economy because it considers both the long run equilibrium relationships and the short run dynamic error corrections at the same time. This is particularly advantageous for economists trying to gather more precise figures for and track the development of the shadow economy and, of course, for the improvement of research in this problematical field of statistics. The preciseness of our estimates also benefit policymakers' efforts to deal with the shadow economy and their formulation of economic policy strategies.

Literature

Aigner, Dennis; Schneider, Friedrich and Damayanti Ghosh (1988): Me and my shadow: estimating the size of the US hidden economy from time series data, in W. A. Barnett; E. R. Berndt and H. White (eds.)*: Dynamic econometric modeling*, Cambridge (Mass.): Cambridge University Press, pp. 224-243.

Bajada, Christopher and Friedrich Schneider (2005): The Shadow Economies of the Asia-Pacific, Pacific Economic Review 10/3, pp. 379-401.

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

Dell' Anno, Roberto, Miguel Gómez-Antonio and Angel Pardo (2007): The shadow economy in three Mediterranean countries: France, Spain and Greece. A MIMIC approach, *Empirical Economics*, 33, pp. 51-84.

Engle, Robert F. and Clive W. J. Granger (1987): Cointegration and Error Correction: Representation, Estimation, and Testing, *Econometrica*, 55/2, March, pp. 251-276.

Engle, Robert F. and Byung Sam Yoo (1987): Forecasting and Testing in Cointegrated Systems, *Journal of Econometrics*, 35, pp. 143-159.

Frey, Bruno S. and Hannelore Weck (1983): Estimating the Shadow Economy: A 'Naive' Approach, *Oxford Economic Papers*, 35, pp. 23-44.

Frey, Bruno S. and Weck-Hannemann, Hannelore (1984): The hidden economy as an "unobserved" variable, *European Economic Review*, 26/1, pp.33-53.

Giles, David, E.A. (1999): Measuring the hidden economy: Implications for econometric modelling, *The Economic Journal*, 109/456, pp. 370-380.

Giles, David, E.A. and Lindsay M. Tedds (2002): *Taxes and the Canadian Underground Economy*, Canadian Tax Paper No. 106, Toronto: Canadian Tax Foundation.

Giles, David, E.A., Tedds, Lindsay, M. and Werkneh, Gugsa (2002): The Canadian underground and measured economies, *Applied Economics*, 34/4, pp. 2347-2352.

Granger, Clive W. J. and A. A. Weiss (1983): Time Series Analysis of Error-Correcting Models, in *Studies in Econometrics, Time Series, and Multivariate Statistics*, New York: Academic Press, pp. 255-278.

Greene W. H. (2007): Econometric Analysis, Prentice-Hall: London.

Johnson, Simon; Kaufmann, Daniel and Pablo Zoido-Lobatón (1998a): Regulatory discretion and the unofficial economy. *The American Economic Review*, 88/ 2, pp. 387-392.

Johnson, Simon; Kaufmann, Daniel and Pablo Zoido-Lobatón (1998b): Corruption, public finances and the unofficial economy. *Discussion paper*, The World Bank, Washington, D.C.

Kwiatkowski, Denis, Peter C. B. Phillips, Peter Schmidt and Yongcheol Shin (1992): *Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root*, Journal of Econometrics, 54, 159-178.

Lippert, Owen and Michael Walker (eds.) (1997): *The Underground Economy: Global Evidences of its Size and Impact*, Vancouver: The Frazer Institute.

MacKinnon, James G. (1996): Numerical Distribution Functions for Unit Root and Cointegration Tests, *Journal of Applied Econometrics*, 11, 601-618.

Mummert, Annette and Friedrich Schneider (2002): The German shadow economy: Parted in a united Germany?, *Finanzarchiv*, 58/3, pp. 286-316.

Newey, Whitney and Kenneth West (1994): Automatic Lag Selection in Covariance Matrix Estimation, *Review of Economic Studies*, 61, 631-653.

Sargan, J. D. (1964): Wages and prices in the United Kingdom: A study in economic methodology, in P. Hart, G. Mills, and J. N. Whittaker (eds.), *Econometric Analysis for National Economic Planning*, Butterworths: London.

Schneider, Friedrich (1994): Can the shadow economy be reduced through major tax reforms? An empirical investigation for Austria, *Supplement to Public Finance/ Finances Publiques*, 49, pp. 137-152.

Schneider, Friedrich (1997): The shadow economies of Western Europe, *Economic Affairs*, 17/3, pp. 42-48.

Schneider, Friedrich (2003): The shadow economy, in: Charles K. Rowley and Friedrich Schneider (eds.), *Encyclopedia of Public Choice*, Dordrecht: Kluwer Academic Publishers.

Schneider, Friedrich (2005): Shadow economies around the world: what do we really know?, *European Journal of Political Economy*, 21(3), September, pp. 598-642.

Schneider, Friedrich and Dominik Enste (2000): Shadow economies: Size, causes, and consequences, *The Journal of Economic Literature*, 38/1, pp. 77-114.

Schneider, Friedrich and Dominik Enste (2002): *The Shadow Economy: Theoretical Approaches, Empirical Studies, and Political Implications*, Cambridge: Cambridge University Press.

Tanzi, Vito (1999): Uses and abuses of estimates of the underground economy, *The Economic Journal*, 109/456, pp. 338-340.

Thomas, Jim J. (1992): *Informal Economic Activity*, LSE, Handbooks in Economics, London: Harvester Wheatsheaf.

Weck, Hannelore (1983): *Schattenwirtschaft: Eine Möglichkeit zur Einschränkung der öffentlichen Verwaltung? Eine ökonomische Analyse*, Frankfurt/Main: Lang.

Appendix A: Deriving the MIMIC Model's Covariance Matrix

The MIMIC model's structural and measurement equations are $\eta_t = \gamma' \mathbf{x_t} + \varsigma_t$ and $\mathbf{y_t} = \lambda \eta_t + \varepsilon_t$, respectively. Expressing this model in terms of covariances gives:

$$\Sigma = \begin{pmatrix} \mathbf{Var(y_t)} & \mathbf{Cov(y_t, x_t)} \\ \mathbf{Cov(x_t, y_t)} & \mathbf{Var(x_t)} \end{pmatrix} = \mathbf{E}\left(\begin{bmatrix} \mathbf{y_t} \\ \mathbf{x_t} \end{bmatrix} \begin{bmatrix} \mathbf{y_t} \\ \mathbf{x_t} \end{bmatrix}' \right).$$

After taking the transposes, multiplications, and making use of the assumptions that:

1. the variables are measured as deviations from mean, i.e. $\mathbf{E(\eta_t) = E(x_t) = E(\varsigma_t) = E(y_t) = E(\varepsilon_t) = 0}$;

2. the error terms do not correlate to the causes, i.e. $\mathbf{E(x_t \varsigma_t') = E(\varsigma_t x_t') = 0}$ and $\mathbf{E(x_t \varepsilon_t') = E(\varepsilon_t x_t') = 0}$;

3. the error terms do not correlate across equations, $\mathbf{E(\varepsilon_t \varsigma_t') = E(\varsigma_t \varepsilon_t') = 0}$; and,

4. the errors of the measurement model do not correlate to the latent variable, i.e. $\mathbf{E(\eta_t \varepsilon_t') = E(\varepsilon_t \eta_t') = 0}$;

we distribute the expectation operator and can thus derive both the variance and covariance between the observable variables. By doing this, it follows that:

$$\begin{aligned}
\mathbf{E(y_t y_t')} &= \mathbf{E}\left[(\lambda \eta_t + \varepsilon_t)(\lambda \eta_t + \varepsilon_t)' \right] \\
&= \mathbf{E}(\lambda \eta_t \eta_t' \lambda' + \lambda \eta_t \varepsilon_t' + \varepsilon_t \eta_t' \lambda' + \varepsilon_t \varepsilon_t') \\
&= \lambda \mathbf{E}(\eta_t \eta_t')\lambda' + \mathbf{\Theta_\varepsilon} \\
&= \lambda \mathbf{E}\left[(\gamma' \mathbf{x_t} + \varsigma_t)(\gamma' \mathbf{x_t} + \varsigma_t)' \right]\lambda' + \mathbf{\Theta_\varepsilon} \\
&= \lambda \mathbf{E}(\gamma' \mathbf{x_t} \mathbf{x_t'} \gamma + \gamma' \mathbf{x_t} \varsigma_t' + \varsigma_t \mathbf{x_t'} \gamma + \varsigma_t \varsigma_t')\lambda' + \mathbf{\Theta_\varepsilon} \\
&= \lambda (\gamma' \mathbf{\Phi} \gamma + \psi)\lambda' + \mathbf{\Theta_\varepsilon}
\end{aligned}$$,

$$\begin{aligned}
\mathbf{E(x_t y_t')} &= \mathbf{E}\left[\mathbf{x_t}(\lambda \eta_t + \varepsilon_t)' \right] \\
&= \mathbf{E}(\mathbf{x_t} \eta_t' \lambda' + \mathbf{x_t} \varepsilon_t' + \varepsilon_t \eta_t' \lambda' + \varepsilon_t \varepsilon_t') \\
&= \mathbf{E}(\mathbf{x_t} \eta_t')\lambda' \\
&= \mathbf{E}\left[(\mathbf{x_t})(\gamma' \mathbf{x_t} + \varsigma_t)' \right]\lambda' \\
&= \mathbf{E}(\mathbf{x_t} \mathbf{x_t'} \gamma + \mathbf{x_t} \varsigma_t')\lambda' \\
&= \mathbf{\Phi} \gamma \lambda'
\end{aligned}$$,

$$\begin{aligned}
\mathbf{E(y_t x_t')} &= (\mathbf{\Phi} \gamma \lambda')' \\
&= \lambda \gamma' \mathbf{\Phi}
\end{aligned}$$,

$$E\left(\mathbf{x_t x_t'}\right) = \mathbf{\Phi} .$$

Thus, $\mathbf{\Theta_\varepsilon}$ is the covariance matrix of the error terms in the measurement model; ψ is the variance of the error term in the structural equation; and, $\mathbf{\Phi}$ is the covariance matrix of the causes. Finally, the covariance matrix of the MIMIC model is:

$$\mathbf{\Sigma} = \begin{pmatrix} \lambda\left(\mathbf{\gamma'\Phi\gamma} + \psi\right)\mathbf{\lambda'} + \mathbf{\Theta_\varepsilon} & \lambda\mathbf{\gamma'\Phi)} \\ \mathbf{\Phi\gamma\lambda'} & \mathbf{\Phi} \end{pmatrix} .$$

Appendix B: Covariance Matrix of the Long Run Part

The EMIMIC model's long run equations with I(0) and I(1) causes are $\mathbf{y_t} = \lambda\mathbf{\eta_t} + \mathbf{\varepsilon_t}$ and $\mathbf{\eta_t} = \mathbf{\gamma'x_t} + \mathbf{\tau'v_t} + \varsigma_t$. As a result, the covariance matrix in its general form is given as:

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{Var(y_t)} & & \\ \mathbf{Cov(x_t, y_t)} & \mathbf{Var(x_t)} & \\ \mathbf{Cov(v_t, y_t)} & \mathbf{Cov(v_t, x_t)} & \mathbf{Var(v_t)} \end{pmatrix} = \mathbf{E}\left(\begin{bmatrix} \mathbf{y_t} \\ \mathbf{x_t} \\ \mathbf{v_t} \end{bmatrix} \begin{bmatrix} \mathbf{y_t} \\ \mathbf{x_t} \\ \mathbf{v_t} \end{bmatrix}' \right) .$$

After multiplication, distributing the expectations operator, and using the necessary assumptions, we get the covariance matrix in terms of model parameters and covariances. Its sub-matrices are:

$$\begin{aligned}
\mathbf{E}\left(\mathbf{y_t y_t'}\right) &= \mathbf{E}\left[\left(\lambda\mathbf{\eta_t} + \mathbf{\varepsilon_t}\right)\left(\lambda\mathbf{\eta_t} + \mathbf{\varepsilon_t}\right)' \right] \\
&= \mathbf{E}\left(\lambda\mathbf{\eta_t \eta_t'\lambda'} + \lambda\mathbf{\eta_t\varepsilon_t'} + \mathbf{\varepsilon_t\eta_t'\lambda'} + \mathbf{\varepsilon_t\varepsilon_t'}\right) \\
&= \lambda\mathbf{E}\left(\mathbf{\eta_t\eta_t'}\right)\mathbf{\lambda'} + \mathbf{\Theta_\varepsilon} \\
&= \lambda\mathbf{E}\left[\left(\mathbf{\gamma'x_t} + \mathbf{\tau'v_t} + \varsigma_t\right)\left(\mathbf{\gamma'x_t} + \mathbf{\tau'v_t} + \varsigma_t\right)' \right]\mathbf{\lambda'} + \mathbf{\Theta_\varepsilon} \\
&= \lambda\mathbf{E}\left(\mathbf{\gamma'x_t x_t'\gamma} + \mathbf{\gamma'x_t v_t'\tau} + \mathbf{\tau'v_t x_t'\gamma} + \mathbf{\tau'v_t v_t'\tau}\right)\mathbf{\lambda'} + \mathbf{\Theta_\varepsilon} \\
&= \lambda\left(\mathbf{\gamma'\Phi_1\gamma} + 2\mathbf{\gamma'N\tau} + \mathbf{\tau'\Phi_2\tau}\right)\mathbf{\lambda'} + \mathbf{\Theta_\varepsilon}
\end{aligned}$$

$$\mathbf{E}\left(\mathbf{x_t x_t'}\right) = \mathbf{\Phi_1}, \quad \mathbf{E}\left(\mathbf{v_t v_t'}\right) = \mathbf{\Phi_2}, \quad \mathbf{E}\left(\mathbf{v_t x_t'}\right) = \mathbf{N'},$$

$$
\begin{aligned}
\mathbf{E}(\mathbf{x_t y_t'}) &= \mathbf{E}\Big[\mathbf{x_t}\big(\lambda\eta_t + \varepsilon_t\big)'\Big] \\
&= \mathbf{E}(\mathbf{x_t}\eta_t'\lambda' + \mathbf{x_t}\varepsilon_t') \\
&= \mathbf{E}(\mathbf{x_t}\eta_t')\lambda' \\
&= \mathbf{E}\Big[(\mathbf{x_t})(\gamma'\mathbf{x_t} + \tau'\mathbf{v_t} + \varsigma_t)'\Big]\lambda' \\
&= \mathbf{E}(\mathbf{x_t}\mathbf{x_t'}\gamma + \mathbf{x_t}\mathbf{v_t'}\tau)\lambda' \\
&= (\mathbf{\Phi_1}\gamma + \mathbf{N}\tau)\lambda'
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{E}(\mathbf{v_t y_t'}) &= \mathbf{E}\Big[\mathbf{v_t}\big(\lambda\eta_t + \varepsilon_t\big)'\Big] \\
&= \mathbf{E}(\mathbf{v_t}\eta_t')\lambda' \\
&= \mathbf{E}\Big[(\mathbf{v_t})(\gamma'\mathbf{x_t} + \tau'\mathbf{v_t} + \varsigma_t)'\Big]\lambda', \\
&= \mathbf{E}(\mathbf{v_t}\mathbf{x_t'}\gamma + \mathbf{v_t}\mathbf{v_t'}\tau)\lambda' \\
&= (\mathbf{N}'\gamma + \mathbf{\Phi_2}\tau)\lambda'
\end{aligned}
$$

where $\mathbf{\Phi_1}$ and $\mathbf{\Phi_2}$ are sub-covariance matrices of the I(0) and I(1) causes, respectively. We further define $\mathbf{N} := \mathbf{E}(\mathbf{v_t}\mathbf{x_t'})$. Finally, we obtain:

$$
\Sigma = \begin{pmatrix}
\lambda(\gamma'\mathbf{\Phi_1}\gamma + 2\gamma'\mathbf{N}\tau + \tau'\mathbf{\Phi_2}\tau)\lambda' + \mathbf{\Theta_\varepsilon} & & \\
(\mathbf{\Phi_1}\gamma + \mathbf{N}\tau)\lambda' & \mathbf{\Phi_1} & \\
(\mathbf{N}'\gamma + \mathbf{\Phi_2}\tau)\lambda' & \mathbf{N}' & \mathbf{\Phi_2}
\end{pmatrix}.
$$

Appendix C: Covariance Matrix of the Short Run Part

Both of the short run part of the EMIMIC model's equations are $\Delta\mathbf{y_t} = \lambda\Delta\eta_t + \varepsilon_t$ and $\Delta\eta_t = \alpha\Delta\mathbf{x_t} + \beta\mathbf{v_t} + \kappa\mathbf{z_{t-1}} + \varsigma_t$. As a result, the model's general covariance matrix is given as:

$$
\Sigma = \begin{pmatrix}
\mathbf{Var}(\Delta\mathbf{y_t}) & & & \\
\mathbf{Cov}(\Delta\mathbf{x_t}, \Delta\mathbf{y_t}) & \mathbf{Var}(\Delta\mathbf{x_t}) & & \\
\mathbf{Cov}(\mathbf{v_t}, \Delta\mathbf{y_t}) & \mathbf{Cov}(\mathbf{v_t}, \Delta\mathbf{x_t}) & \mathbf{Var}(\mathbf{v_t}) & \\
\mathbf{Cov}(\mathbf{z_{t-1}}, \Delta\mathbf{y_t}) & \mathbf{Cov}(\mathbf{z_{t-1}}, \Delta\mathbf{x_t}) & \mathbf{Cov}(\mathbf{z_{t-1}}, \mathbf{v_t}) & \mathbf{Var}(\mathbf{z_{t-1}})
\end{pmatrix} = \mathbf{E}\left(\begin{bmatrix} \Delta\mathbf{y_t} \\ \Delta\mathbf{x_t} \\ \mathbf{v_t} \\ \mathbf{z_{t-1}} \end{bmatrix}\begin{bmatrix} \Delta\mathbf{y_t} \\ \Delta\mathbf{x_t} \\ \mathbf{v_t} \\ \mathbf{z_{t-1}} \end{bmatrix}'\right).
$$

In terms of model parameters and covariances, it becomes:

$$\Sigma = \begin{pmatrix} \lambda(\alpha'\Phi_3\alpha + 2\alpha'M\beta + \beta'\Phi_2\beta + \kappa'\Omega\kappa)\lambda' + \Theta_\varepsilon & & & \\ (\Phi_3\alpha + M'\beta)\lambda' & \Phi_3 & & \\ (M\alpha + \Phi_2\beta)\lambda' & M & \Phi_2 & \\ \lambda(\psi\lambda'\kappa + \psi)\lambda' + \Theta_\varepsilon & 0 & 0 & \Omega \end{pmatrix}.$$

In addition to previous definitions, we define $M := E(v_t, \Delta x_t')$, $\Phi_3 := E(\Delta x_t \Delta x_t')$, and assume that the long run part's error term is a stationary white noise series. Consequently, its covariance matrix should not change with time, i.e. $\text{Cov}(z_t) = \text{Cov}(z_{t-1}) = \Omega$. We further assume that $E(x_t \varsigma_{t-1}') = 0$, $E(\varsigma_{t-1} x_t') = 0$, $E(x_t \varepsilon_{t-1}') = 0$, and $E(\varepsilon_{t-1} x_t') = 0$. In addition, $E(\eta_t \varepsilon_{t-1}') = 0$ and $E(\varepsilon_{t-1} \eta_t') = 0$. As a result, the sub-matrices are derived as follows:

$$\begin{aligned} E(\Delta y_t \Delta y_t') &= E\left[(\lambda\Delta\eta_t + \varepsilon_t)(\lambda\Delta\eta_t + \varepsilon_t)' \right] \\ &= E(\lambda\Delta\eta_t\Delta\eta_t'\lambda' + \lambda\Delta\eta_t\varepsilon_t' + \varepsilon_t\Delta\eta_t'\lambda' + \varepsilon_t\varepsilon_t') \\ &= \lambda E(\eta_t\eta_t')\lambda' + \Theta_\varepsilon \\ &= \lambda E\left[(\alpha'\Delta x_t + \beta'v_t + \kappa'z_{t-1} + \varsigma_t)(\alpha'\Delta x_t + \beta'v_t + \kappa'z_{t-1} + \varsigma_t)' \right]\lambda' + \Theta_\varepsilon , \\ &= \lambda[\alpha'\Phi_3\alpha + \alpha'M\beta + \alpha'E(\Delta x_t z_{t-1}')\kappa + \beta'M'\alpha + \beta'\Phi_2\beta + \\ &\quad + \beta'E(v_t z_{t-1}')\kappa + \kappa'E(z_{t-1}\Delta x_t')\alpha + \kappa'E(z_{t-1}v_t')\beta + \kappa'\Omega\kappa]\lambda' + \Theta_\varepsilon \\ &= \lambda(\alpha'\Phi_3\alpha + 2\alpha'M\beta + \beta'\Phi_2\beta + \kappa'\Phi_2\kappa)\lambda' + \Theta_\varepsilon \end{aligned}$$

$$\begin{aligned} E(\Delta x_t y_t') &= E\left[\Delta x_t (\lambda\Delta\eta_t + \varepsilon_t)' \right] \\ &= E(\Delta x_t \Delta\eta_t')\lambda' \\ &= E\left[(\Delta x_t)(\alpha'\Delta x_t + \beta'v_t + \kappa'z_{t-1} + \varsigma_t)' \right]\lambda' , \\ &= (\Phi_3\alpha + M'\beta + E[(\Delta x_t)(\lambda\varsigma_{t-1} + \varepsilon_{t-1})]\kappa)\lambda' \\ &= (\Phi_3\alpha + M'\beta)\lambda' \end{aligned}$$

$$\begin{aligned} E(v_t \Delta y_t') &= E\left[v_t (\lambda\Delta\eta_t + \varepsilon_t)' \right] \\ &= E\left[(v_t)(\alpha'\Delta x_t + \beta'v_t + \kappa'z_{t-1} + \varsigma_t)' \right]\lambda' , \\ &= E(v_t\Delta x_t'\alpha + v_t v_t'\beta + v_t z_{t-1}'\kappa)\lambda' \\ &= (M\alpha + \Phi_2\beta)\lambda' \end{aligned}$$

$$\begin{aligned}
\mathbf{E}(z_{t-1}\Delta y'_t) &= \mathbf{E}\left[\left(\lambda\varsigma_{t-1} + \varepsilon_{t-1}\right)\left(\lambda\Delta\eta_t + \varepsilon_t\right)'\right] \\
&= \lambda\mathbf{E}\left(\varsigma_{t-1}\Delta\eta'_t\right)\lambda' + \mathbf{\Theta_\varepsilon} \\
&= \lambda\mathbf{E}\left[\left(\varsigma_{t-1}\right)\left(\alpha'\Delta\mathbf{x_t} + \beta'\mathbf{v_t} + \kappa'\mathbf{z_{t-1}} + \varsigma_t\right)'\right]\lambda' + \mathbf{\Theta_\varepsilon} \ , \\
&= \lambda\mathbf{E}\left(\left(\varsigma_{t-1}\right)\left(\lambda\varsigma_{t-1} + \varepsilon_{t-1}\right)'\kappa + \psi\right)\lambda' + \mathbf{\Theta_\varepsilon} \\
&= \lambda\left(\psi\lambda'\kappa + \psi\right)\lambda' + \mathbf{\Theta_\varepsilon} \ ,
\end{aligned}$$

$$\mathbf{E}(\mathbf{z_{t-1}}\Delta\mathbf{x'_t}) = \mathbf{E}\left[\left(\lambda\varsigma_{t-1} + \varepsilon_{t-1}\right)\left(\Delta\mathbf{x_t}\right)'\right] = 0 \ ,$$

$$\mathbf{E}(\mathbf{z_{t-1}}\mathbf{v'_t}) = \mathbf{E}\left[\left(\lambda\varsigma_{t-1} + \varepsilon_{t-1}\right)\left(\mathbf{v_t}\right)'\right] = 0 \ .$$

Appendix D: Data Sources and Additional Test Statistics

Table D.1: *Data Sources*

Variable	Description	Source	Annotations
Causes			
Tax	Tax and Social Security Contributions Burden / GDP	OECD Economic Outlook	Quarterly Data (Q1 1981:Q4 2006), seasonally adjusted
Unemp	Unemployment Rate	OECD Economic Outlook	Quarterly Data (Q1 1981:Q4 2006), seasonally adjusted
Reg	Government Employment / Labor Force	OECD Economic Outlook	Quarterly Data (Q1 1981:Q4 2006), seasonally adjusted
Work	Hours Worked per Employee in Total Economy	OECD Economic Outlook	Index Series (2000=100), Quarterly Data (Q1 1981:Q4 2006), seasonally adjusted
Indicators			
M1	Monetary Aggregate M1	Banque de France	Natural Logarithm, Quarterly Data (Q1 1981:Q4 2006)
GDP	GDP	OECD Economic Outlook	Volume Index (2000=100), Quarterly Data (Q1 1981:Q4 2006), seasonally adjusted

Table D.2: *Further Summary Statistics of the MIMIC Models*

	Long Run MIMIC Model	Short Run MIMIC Model
P-value for Test of Close Fit (RMSEA < 0.05)	1.0000	0.9800
Root Mean Square Residual (RMR)	0.0035	0.0540
Standardized RMR	0.0035	0.0550
Goodness of Fit Index (GFI)	0.99	0.96
Adjusted Goodness of Fit Index (AGFI)	0.99	0.94
Parsimony Goodness of Fit Index (PGFI)	0.61	0.69

Table D.3: *Size of the French Shadow Economy in % of Official GDP*

Year	Q1	Q2	Q3	Q4
1982	12.88	12.86	13.11	13.06
1983	12.96	13.00	13.08	13.20
1984	13.30	13.47	13.32	13.47
1985	13.28	13.28	13.27	13.27
1986	13.39	13.29	13.39	13.58
1987	13.68	13.52	13.60	13.41
1988	13.29	13.40	13.32	13.51
1989	13.57	13.83	13.82	13.62
1990	13.81	13.98	13.99	14.14
1991	13.89	13.99	13.99	13.97
1992	13.71	13.98	14.01	14.13
1993	13.99	14.06	14.04	14.07
1994	14.03	13.96	14.02	14.12
1995	14.30	14.23	14.58	14.66
1996	14.58	14.49	14.54	14.52
1997	14.34	14.30	14.28	14.31
1998	14.33	14.35	14.56	14.68
1999	14.66	14.64	14.69	14.84
2000	14.83	15.09	15.27	15.25
2001	15.31	15.48	15.45	15.69
2002	15.52	15.68	15.77	15.89
2003	15.73	15.82	15.61	15.57
2004	15.52	15.35	15.42	15.34
2005	15.43	15.62	15.56	15.76
2006	15.72	15.72	15.93	15.93

Table D.4: *EMIMIC vs. Long Run Equilibrium (% of Official GDP)*

Year	EMIMIC	Long Run Equilibrium Relationship	Deviation from Long Run Equilibrium
1982	13.06	12.99	0.06
1983	13.20	13.03	0.18
1984	13.47	13.34	0.13
1985	13.27	13.36	-0.10
1986	13.58	13.49	0.09
1987	13.41	13.56	-0.15
1988	13.51	13.57	-0.06
1989	13.62	13.83	-0.21
1990	14.14	14.03	0.11
1991	13.97	13.93	0.04
1992	14.13	13.96	0.17
1993	14.07	13.96	0.12
1994	14.12	14.19	-0.07
1995	14.66	14.48	0.18
1996	14.52	14.47	0.05
1997	14.31	14.43	-0.12
1998	14.68	14.62	0.06
1999	14.84	14.97	-0.13
2000	15.25	15.35	-0.10
2001	15.69	15.55	0.15
2002	15.89	15.76	0.14
2003	15.57	15.53	0.04
2004	15.34	15.43	-0.09
2005	15.76	15.78	-0.03
2006	15.93	16.01	-0.09