A note on the estimation of an equation over a panel of countries[1]

Jean-Pierre LAFFARGUE[2]

CEPREMAP and TEAM

Abstract: Estimating a macroeconomic equation over a panel of countries by GMM has become popular. However, the covariance and autocovariance matrices of the shocks hitting the countries at a same time have a large size and are estimated on a rather short time period. To improve the precision of their estimation, we assume that the structure of the shocks can be represented by a limited number of common factors and we apply recent developments of factor analysis.

JEL Codes: C23

Keywords: Econometrics of panels, Covariance matrix, Factor analysis, Gennnneralizes method of moments.

Introduction

Estimating a macroeconomic equation over a panel of countries has become popular. According to this approach the values of the parameters of this equation, but not its specification, may differ or not between countries. Using panel estimation helps to get more robust and precise empirical findings: as these countries share some common structural features, each country estimation benefits from information brought by its partners. Moreover, panel estimation allows to identify deep structural differences between countries.

As the errors terms of the various countries are probably correlated, and as the structure of these correlations are probably more complex than the one allowed by error components models, SUR methods appear the most natural way to make this estimation. However, the presence of endogenous and anticipated explanatory variables requires the use of instrumental variables and GMM methods, instead of generalized least squares. In both cases, the covariance matrix of the shocks hitting the countries at a same time has a large dimension and is estimated on a rather short time period. This problem is made more complex if we allow for some autocorrelation between shocks. To improve the precision of the estimation of the covariance and the autocovariances of the shocks and to solve the problems which might result from their singularity, we assume that the structure of the shocks can be represented by a limited number of common factors and we use recent developments of factor analysis.

The problem

We want to estimate on a panel of I countries, indexed by i, and on a period of T years, indexed by t, the following system of I equations:

(1) $y_{it} = g(y_{i,t-1}, x^a_{1i,t+1}, x_{2it}, x_{3it}; \alpha_i) + \varepsilon_{it}; i = 1,..,I; t = 2,..,T-1$.

[1] The author thanks Stéphanie Guichard, Philippe Jolivaldt and Patrick Sevestre for their remarks and suggestions. The remaining imperfections of this paper are of my responsibility.

[2]Cepremap, 142 rue du Chevaleret, 75013 PARIS France. Tél 33 1 40 77 84 26. E-mail jplaffarg1@cybercable.fr

I and T are of the same order of magnitude, and not very high. The y_{it} are the explained variables, the x_{jit} are the explanatory variables, g is a function representing the behavior associated to country i, the α_i are the parameters of this function (they may differ or not between countries). ε_{it} is the error term of null expected value[3].

We assume that the error terms of a common year are correlated, and call Ω, of typical element ω_{ij}, their covariance matrix. This assumption is consistent with an interpretation of error terms as correlated random shocks affecting the different domestic economies. We will put some structure on these correlations to increase the number of degrees of freedom of our estimation. However, the structure usually proposed by error component models is too restrictive for our needs. We will assume either that the error terms are non autocorrelated or that they are autocorrelated.

If all the explanatory variables were predetermined, that is if the ε_{it} were independent of the contemporaneous and past values of the explanatory variables, system (1) could be easily estimated by generalized nonlinear least squares. However, we prefer making more general assumptions. Thus, we assume that variable x_{2it} is predetermined, but that this property is not shared by variable x_{3it}. Moreover, variable $x^{a}_{1i,t+1}$ represents the forecast at time t of variable x_{1i} for time $t+1$. As this variable is not observed, we follow a suggestion by Wickens (1981), and substitute it by its observed value at time $t+1$: $x_{1i,t+1}$. Thus, we introduce a supplementary error in the equation, which bears on the foreseen value of an explanatory variable for a future time[4]. We have then to estimate a model with errors on variables. The endogeneity of some variables and the error in some others make least squares estimators non consistent.

To overcome these difficulties we allocate to each national equation a vector line of n instrumental variables W_{it} and we assume that the processes they follow are uncorrelated with the processes of the ε_{it}.. Then, we will use a two steps GMM method.

Estimation of the system of equations (1) by GMM

Let W_i be the matrix of observations for the instruments related to country i, of size $(T-2, n)$. W_{it} is its typical line. Then, we define by $V_t = (\varepsilon_{1t}W_{1t} .... \varepsilon_{It}W_{It})$ the line vector of size In, and by V the matrix with typical line V_t and dimension $(T-2, In)$. The moment's condition is:

(2) $EV_t = 0$

We approximate the theoretical moments by the empirical moments and we get:

(3) $V'\iota = 0$

where ι is a column vector of 1 with dimension $T-2$. Condition (3) cannot be exactly checked in most cases where the total number of instruments is larger than the number of parameters to

[3] Our problem is different from the one of Pesaran and Smith (1995). We directly estimate the various values that a given parameter can take in the various countries without supplementary assumptions. For Pesaran and Smith, the differences between these values are random, and they estimate the expected value and the variance of each coefficient over the set of all countries. They could also compute in their analytic frame the optimal forecast of the values that the coefficients take in the various countries. Thus, if we use the terminology of the econometrics of panels, our approach is similar to that of the models with fixed-effects, and the approach of Pesaran and Smith is similar to the one of component errors models. This last approach gives more precise estimations, if the stronger assumptions it requires are valid.

[4] And then which follows a difference of martingale, independent of the explanatory and explained variables at and before current time.

estimate. Thus, we try to minimize the distance between $V'\iota$ and 0, by using a distance matrix A, of dimension: (In, In), which is symmetric and positive definite. Thus, we minimize relatively to parameters the expression:

(4) $\iota' VAV' i$

The efficient choice of matrix A is: $A = \Phi^{-1}$, where Φ. is the spectral density at frequency 0 of $V'\iota/(T-2)^{1/2}$. We can separate two cases. When the process of the error term is non autocorrelated, the estimation of Φ has for typical element: $\omega_{ij} W_i^{'} W_j /(T-2)$. Then, to compute A, we must invert this matrix, of dimension (In, In)[5].

When the process of the error term is autocorrelated we note $\Gamma(h) = E\varepsilon_t \varepsilon_{t+h}^{'}, \forall h \in Z$, its autocovariance function, with typical element $\gamma_{ij}(h)$. Let us choose a kernel $\kappa(\ )$ and a bandwith parameter ξ. . Then, the estimation of Φ: has for typical element: $[\kappa(0)\omega_{ij} W_i^{'} W_j + \sum_{h=1}^{T-3} \kappa(h/\xi_{T-2})\gamma_{ij}(h) \sum_{t=2}^{T-1-h} W_{it}^{'} W_{j,t+h}]/(T-2)$.

Den Haan and Levin (1996) is a good guide for the choice of the kernel and the bandwith.

In practice we proceed through two steps. In the first step, we assume the errors terms to be non autocorrelated and with a covariance matrix Ω proportional to the identity matrix. Thus, A is the block diagonal matrix, with typical block: $(W_i^{'} W_i)^{-1}$. We minimize criteria (4), and thus we get a first value for the parameters and the residuals. Then, we can compute estimators of the covariance and the autocovariance matrices of the error terms and, in the second step, apply the previous formulae. This second step may be iterated several times.

The covariance matrix of the estimated parameters time $(T-2)^{1/2}$, is asymptotically equal to $(\Delta' \Phi^{-1} \Delta)^{-1}$, where Δ is the matrix of the partial derivatives of $V'\iota/(T-2)$ relatively to the parameters.

A difficulty is that the estimation of the covariance and autocovariance matrices of the error terms is very imprecise: $\hat{\varepsilon}_{it}$ is observed for $t = 2,..,T-1$, which makes $T-2$ observations. Yet I is of the order of $T-2$. Thus these matrices are almost singular, or even singular if the number of observed years is smaller than the number of countries. We use factor analysis to put some structure in this matrix; that is some interdependence between the shocks hitting the countries in a way that should appear natural to economists.

[5] We have assumed that the covariance matrix Ω, is independent of time (time homoscedasticity), In the computation of Φ we could expect to get estimators of the parameters and of their covariance matrix robust to time-heteroscedasticity by substituting in the expression of Φ, $\hat{\omega}_{ij} W_i^{'} W_j /(T-2)$ by $\sum_{t=2}^{T-1} W_{it}^{'} \hat{\varepsilon}_{it} \hat{\varepsilon}_{jt} W_{jt} /(T-2)$. The problem is that this new estimator is the sum of $T-2$ matrices of dimensions (In, In), but of rank 1. Indeed the term of time t is the product of the column vector $V_t^{'}$ by its transpose. Consequently, the rank of the estimator of Φ is at most equal to $T-2$. In most applications it will be less than In, and matrix Φ will be singular so non-invertible. Thus, it seems impossible to build estimators robust to heteroscedasticity for our problem. For the same reason we will assume the autocovariance matrices of the error terms to be independent of time.

Estimation of the covariance matrix of the error terms when they are non autocorrelated[6]

ε_t denotes the vector of error terms for the set of all nations (of dimension I) and for $t = 2,..,T-1$. We denote in the same way the random vector, its realization and its estimation. We make the following assumptions:

(5) $\varepsilon_t = \Lambda F_t + u_t$,

F_t represents a column vector of dimension f; its elements are called common factors. u_t is a column vector of dimension I; its elements are called specific components. Both are random. Λ is a matrix of dimension (I, f) and is certain. Its elements are called loadings.

$EF_t = Eu_t = 0$, $E(u_t u_t^{'}) = D = diag(d_1,..,d_I)$[7], $E(F_t u_\tau^{'}) = 0$, $\forall t, \tau$.

$E(F_t F_\tau^{'}) = E(u_t u_\tau^{'}) = 0, \forall \tau, t \neq \tau$, $E(F_t F_t^{'}) = U_f$[8].

Then, we deduce:

(6) $\Omega = \Lambda\Lambda' + D$

Instead of having to estimate the $I(I+1)/2$ parameters of Ω, we just have to estimate the $(f+1)I$ parameters of Λ and D (actually the improvement is meaningful only when the number of factors is much smaller than half the number of countries). It can be shown that the maximum likelihood estimators of Λ and D, denoted by $\hat{\Lambda}$ and $\hat{D}$, under the assumption of normality of ε_t, are given by conditions:

$\hat{\Omega} = \sum_{t=2}^{T-1} (\varepsilon_t - m)(\varepsilon_t - m)'/(T-2)$, where m is the arithmetic mean vector of the ε_t over the estimation period.

$1+\gamma_1,..,1+\gamma_I$, are the real positive eigenvalues of $\hat{\Omega}\hat{D}^{-1}$, which are assumed to be different and ranked by decreasing values (actually, the f first γ_i must be positive for the computation to be possible),

Γ is the diagonal matrix of dimension (f, f) with diagonal elements: $\gamma_1, .\ \gamma_f$.

the f columns of $\hat{\Lambda}$ are the f first eigen vectors of $\hat{\Omega}\hat{D}^{-1}$ (related to the f largest eigenvalues) which are normed to check for the identification condition: $\Gamma = \hat{\Lambda}\hat{D}^{-1}\hat{\Lambda}$.

The estimation procedure is iterative. First, we give an initial value to $\hat{D}$: D_0. Then we compute the eigenvalues and the eigen vectors of $\hat{\Omega}D_0^{-1}$,and consequently Λ_0. Then, we compute D_1 which is the diagonal matrix, the diagonal elements of which are the same as for $\hat{\Omega} - \Lambda_0\Lambda_0^{'}$, and we start again. This procedure appears to converge easily in applications, although to our knowledge there

[6] Doz (1998, page 85-161) gives a clear and rigorous introduction to factor analysis in the case of non autocorrelation, and we base on it here. Doz borrows much from Lawley and Maxwell (1971) and from Bartholomew (1987).

[7] $diag$ means a diagonal matrix with the following diagonal elements.

[8] U_f represents an identity matrix, which in this case is of dimension (f, f).

do not exist mathematical results proving this property. More sophisticated estimation methods exist and are given by Doz[9].

The choice of the initial value D_0 is a supplementary problem. We denote by R_i^2 the square of the multiple correlation coefficient between the *ith* component of ε_t and the $I-1$ other components, and by $\hat{\omega}_{ij}$ the typical element of matrix $\hat{\Omega}$. Then, we choose: $d_{i0} = \hat{\omega}_{ii}(1-R_i^2)$.

Another difficulty is the choice of the number of factors f. A simple method is to compute a matrix of the same dimension as $\hat{\Omega}$, the non diagonal terms of which represent the correlations between the components of vector ε_t, and the diagonal terms of which are the R_i^2. Then we make a principal component analysis of this matrix, and we keep as many factors as there exists non-negligible positive eigenvalues.

This *a priori* test is sufficient at the beginning of a succession of iterations of GMM, when the fact that matrix Φ may be a little wrong bears no serious consequences. However, an *a posteriori* test of the validity of the choice of the number of factors, more rigorous, must be made at the last step of GMM. This test, of the likelihood ratio kind, uses as null hypothesis that the number of factors is equal to f. The alternative hypothesis is that there does not exist any constraint on the covariance matrix Ω. The statistics of the test is:

(7) $\xi = -(T-2)\sum_{j=p+1}^{I} \ln(1+\gamma_j)$.

This statistics asymptotically verifies a χ^2 with a number of degrees of freedom equal to $[(I-p)^2-(I+p)]/2$. Bartlett suggests substituting, in the expression of ξ, the number of observations: $T-2$, by: $T-2-(2I+5)-2p/3$, when the number of observations is low, which is the situation we face here.

Estimation of the covariance and autocovariance matrices of the error terms when they are autocorrelated

We assume now that each common factor and each specific component follows a weakly stationary process, and may present autocorrelation. We do not make any normality assumption. All the other assumptions of the previous section are kept unchanged. In particular, the factors and the specific components are non correlated to one another. Under these assumptions, the estimator of the last section is a M-estimator. Doz and Lenglart (1999) show that it is consistent. This estimator is non efficient, and does not give any information on the stochastic process followed by the factors and the specific components. However, the computation of this estimator allows the computation of a pseudo-score test of the number of common factors, developed by Doz and Lenglart[10]. Let us make, as in previous section, the null hypothesis that the covariance of the error terms can be represented by a model with f factors. The alternative hypothesis is that there does not exist any constraint on the covariance of process ε_t. We first have to introduce some new notations. We will represent by index 0 the true value of a parameter and by a ^ its M-estimated value. We call:

[9] The empirical covariance $\hat{\Omega}$ and its estimated approximation $\hat{\Lambda}\hat{\Lambda}'+\hat{D}$ have the same diagonal. This results from the fact that the factor representation does not change variances, but simplifies the structure of the covariances by assuming that it results from a small number of common factors.

[10] Doz (1998) shows that when ε_t is non autocorrelated, the likelihood ratio test of the previous section is asymptotically equivalent to the pseudo-score test that we are going to present.

$\theta = \begin{pmatrix} vec\Lambda \\ d \end{pmatrix}$, with: $d = (d_1,..d_I)$.

$h(\theta)$ is the application which associates to θ the vector: $vech(\Lambda\Lambda'+D)$.

E_I is the duplication matrix of order I with dimension $(I^2, I(I+1)/2)$, which verifies for all symmetric matrix M of dimension I : $vecM = E_I vechM$.

$E_I^+ = (E_I^{'} E_I)^{-1} E_I^{'}$ is the pseudo-inverse of E. More generally, exponent – means pseudo-inverse and exponent + means generalized inverse.

We will assume now that ε_t follows a Gaussian stationary process, and we note: $\forall h \in Z$, $\Gamma(h) = E(\varepsilon_t \varepsilon_{t+h}^{'})$, $B_0 = \sum_{h\in Z} \Gamma(h) \otimes \Gamma(h)$.

$J_0 = D_I^{'} \Gamma(0)^{-1} \otimes \Gamma(0)^{-1} D_I / 2$.

$$P = \frac{\partial h}{\partial \theta'}(\theta_0)\left(\frac{\partial h'}{\partial \theta}(\theta_0) J_0 \frac{\partial h}{\partial \theta}(\theta_0)\right)^{+} \frac{\partial h'}{\partial \theta}(\theta_0) J_0,\ M = U_{I(I+1)/2} - P.$$

Then, the statistics of the pseudo-score test is:

$$\xi = [(T-2)/2] vech(\Lambda\Lambda'+D-\hat{\Omega})'(\hat{M} E_I^{+} B_0 E_I^{+'} \hat{M}')^{-} vech(\Lambda\Lambda'+D-\Omega).$$

This statistics asymptotically verifies a χ^2 with a number of degrees of freedom equal to $[(I-p)^2 - (I+p)]/2$.

Practically, B_0 is computed as a sum including the $\Gamma(h)$ which are different enough from 0. We can also check that increasing the range of the sum has no significant effect on the numerical value of the test.

When we have determined the number of factors, we can estimate the model of the error terms by computing the likelihood function with a Kalman filter as Stock and Watson (1993) suggested.. Let us assume to simplify the presentation that each common factor follows an ARMA(1,1) and that each specific component follows an AR(1). The model can be written:

(8) $\varepsilon_t = \Lambda F_t + u_t$

$F_t = \varphi F_{t-1} + \nu_t - \theta \nu_{t-1}$

$u_t = \rho u_{t-1} + \zeta_t$

φ and ρ are diagonal matrices of respective dimensions f and I. ν_t and ζ_t are random vectors of dimensions f and I. Their components are non correlated to one another and non autocorrelated. Moreover, we will assume that they are Gaussian. The model can be rewritten as an inobservable component model in a state-measure form:

(9) $\varepsilon_t = \begin{pmatrix} \Lambda & 0 & U_I \end{pmatrix} \begin{pmatrix} F_t \\ \nu_t \\ u_t \end{pmatrix}$

$$\begin{pmatrix} F_t \\ \nu_t \\ u_t \end{pmatrix} = \begin{pmatrix} \varphi & -\theta & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \rho \end{pmatrix} \begin{pmatrix} F_{t-1} \\ \nu_{t-1} \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} U_p & 0 \\ U_p & 0 \\ 0 & U_I \end{pmatrix} \begin{pmatrix} \nu_t \\ \zeta_t \end{pmatrix},$$

or, with an evident change in notations:

(10) $\varepsilon_t = Z\alpha_t$

$\alpha_t = C\alpha_{t-1} + \eta_t$

ε_t is the measure vector which is observable. α_t is the state vector. η_t is the perturbation vector, which is non autocorrelated, and the components of which are nor correlated to one another. We want to estimate elements of matrices Z and C and the variances of the perturbations. This can be easily done by maximum likelihood, using optimal forecasts .of the state vector and the covariance matrices of the forecast error, which are given by the use of Kalman filter (see Harvey (1989)).

A practical problem is that we must choose the number of lags of the ARMA processes which fit the best the dynamics of the common factors and of the specific components. The only solution we see is an error an trial method, which checks for the significativity of the estimated parameters and the quality of the innovations η_t .

When the estimation has been made, it is easy to derive the autocovariance matrices of ε_t , $\Gamma(h)$, under the assumption that model (8) is valid.

Conclusion

We have applied the previous methodology to the estimation of a wage curve over a sample of 16 industrialized countries, using yearly data from 1982 to 1997 (Guichard and Laffargue (2000))[11]. We limited ourselves to the case where the error terms are non autocorrelated, a test which has not been presented here not rejecting this hypothesis. The specification of the wage equation is the same for all countries, but the values of the parameters may differ. We have retained six instruments per country. We have used a strategy of nested tests to evaluate which parameters are identical between countries and which parameters differ.

We show that wages contracts are fairly long (which implies some nominal rigidity) and that price expectations are quite static. The wedge has a positive but small effect on wages: an increase in social contribution mainly results in smaller earnings for the workers. The employment rate is a better indicator of labor market tensions than the unemployment rate. The elasticity of the wage cost to the employment rate clearly differs across countries. However our results are not inconsistent with Blanchflower and Oswald (1994) findings of an elasticity of the wages to the unemployment rate of around 0.1. Lastly, in some countries, the wage behavior is clearly at odds with our specification, because of very specific labor market institutions (Spain) or because the country has experienced a strong economic shock (Finland). The best model was a model where both the productivity and employment rate parameters are country specific, but where all the other parameters took a common value across countries.

References

Bartholomew D.J. (1987), *Latent Variable Models and Factor Analysis*, Griffin.

Blanchflower D. and A. Oswald (1994), *The Wage Curve*, The MIT Press.

[11] The program was written in TSP 4.4 and is available under request.

Doz C. (1998), "Econométrie des modèles à facteurs dynamiques et exemples d'application en macroéconomie", Ph.D. dissertation in economics, University Paris IX Dauphine, May 12.

Doz C. and F. Lenglart (1999), "Analyse factorielle dynamique : test du nombre de facteurs, estimation et application à l'enquête de conjoncture dans l'industrie", *Annales d'Economie et de Statistique*, n° 54, pp. 91-127.

Guichard Stéphanie and Jean-Pierre Laffargue (2000), “Comparaison de la formation des salaires dans un panel de pays industrialisés », *Economie et Prévision*, fothcoming.

den Haan W.J. and A.T. Levin (1996), "A Practitioner's Guide to Robust Covariance Matrix Estimation", NBER Technical Working Paper, n° 197.

Harvey A. C. (1989), *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Lawley D.N. and A.E. Maxwell (1971), *Factor Analysis as a Statistical Method*, American Elsevier, New York.

Pesaran H. and R. Smith (1995), "Estimating Long-Run Relationships from Heterogeneous Panels", *Journal of Econometrics,* 68, p. 79-113.

Stock J.H. and M.W. Watson (1993), "A Procedure for Predicting Recessions with leading Indicators: Econometric Issues and Recent Experiences" in *Business Cycles, Indicators and Forecasting*, J.H. Stock and M.W. Watson ed., University of Chicago Press.

Wickens W. R. (1982), "The Efficient estimation of Econometric Models with Rational Expectations", *Review of Economic Studies* 49, p. 55-67.