Statement of the estimation problem
Larry Karp, September 29, 2021

1 Introduction

I am trying to estimate a model that seems almost standard. I tried two estimation approaches (Section 4) that return nonsensical and mutually contradictory conclusions. I have relied on two GSIs to code the two methods. It is possible that they both made a coding mistake, but that seems unlikely.

Before attempting a third method (Section 3) I would appreciate the advice of a statistician/econometrician regarding that method. If time permits, it would be great to get insight into why the previous two approaches failed, but I realize that this may be asking too much.

2 The model

The model uses panel data, which when stacked can be written in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \text{ with } \boldsymbol{\varepsilon} \sim \left(\mathbf{0}, \sigma^2\mathbf{V}\right). \tag{1}$$

I can write $\mathbf{V}$ in closed form as a function of two parameters, ρ, λ. The goal is to estimate $\rho, \lambda, \sigma, \boldsymbol{\beta}$. Conditional on ρ, λ, estimation of σ, $\boldsymbol{\beta}$ is straightforward, so the difficulty is in estimating ρ, λ.

The data consists of a panel on emissions (CO2) in region i at time t

$$\begin{aligned} y_{it} &= b_{0i} + h(t) + \nu_{it} \\ \text{with } \nu_{it} &= \rho\nu_{t-1} + \alpha_t + \mu_{it} \end{aligned} \tag{2}$$

and

$$\alpha_t \sim iid\left(0, \sigma_\alpha^2\right), \mu_{i,t} \sim iid\left(0, \sigma_\mu^2\right), \text{ and } \mathbf{E}\left(\alpha_t\mu_{i,\tau}\right) = 0\forall t, \tau.$$

The average shock in period t is

$$\nu_t \equiv \frac{\sum_i \nu_{it}}{n} = \rho\nu_{t-1} + \alpha_t + \theta_t \text{ with } \theta_t \equiv \frac{\sum_i \mu_{it}}{n}.$$

I write this as

$$\nu_t = \rho\nu_{t-1} + \eta_t \text{ with } \eta_t \equiv \alpha_t + \theta_t.$$

Region i's emissions in period t equal a region-specific constant, b_{0i}, a common time trend, $h(t)$, and a shock ν_{it}. This region-time shock, $\nu_{it} = \rho\nu_{t-1} + \alpha_t + \mu_{it}$, consists of a systemic component, $\rho\nu_{t-1} + \alpha_t$, common to all regions, and an idiosyncratic component, μ_{it}. The systemic component is AR(1) and the innovation is the sum of the systemic shock, α_t, and the average of the region-time shocks, $\frac{1}{n}\sum_i \mu_{it}$.[1]

I have a closed form expression for the covariance matrix $\mathbf{E}\nu_{i,t}\nu_{j,t+s}$ (Section 5) in terms of ρ, $\sigma^2 \equiv \sigma_\alpha^2$ and $\lambda \equiv \frac{\sigma_\mu^2}{\sigma_\alpha^2}$. This formula produces $\sigma^2\mathbf{V}$ used in equation 1. The β's for this model consist of only the region-specific constants and the common time trend, e.g. a polynomial. I can estimate these (and then estimate σ^2) using GLS given an estimate of $\mathbf{V}$. This requires estimates of ρ and λ, the parameters I am chiefly interested in.

3 Proposed estimation strategy

Denote the $n \times 1$ column vector of regional observations in year t as $\mathbf{y}_t = (y_{1t}, y_{2t}...y_{nt})'$. Stack the model so that the $nT \times 1$ vector of observations is $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2'...\mathbf{y}_T')'$. The covariance matrix is $\sigma^2\mathbf{V}(\rho, \lambda)$. The m, p element of $\mathbf{V}$ is $v_{m,p}$. The indices m, p depend on i, j, t, s, where s is the absolute value of the difference between time indices corresponding to two observations. Given the indices of $(y_{i,t}, y_{j,t+s})$ I can identify the m, p element of $\mathbf{V}$.

Using the formula for the covariance (Section 5), I have an expression for the m, p element of $\mathbf{V}$. Given the mapping between i, j, t, s and m, p, I therefore have a formula for the covariance of $(y_{i,t}, y_{j,t+s})$ denoted $\sigma^2 v_{i,j,t,s}(\rho, \lambda)$. This expression is a linear function of λ and σ^2 and a nonlinear function of ρ. (Section 5 has an example.)

I have two ideas based on the method of moments. The first seems more plausible.

3.1 First method of moments estimator

The algorithm begins with a guess of ρ, λ, thus generating a guess of the covariance matrix $\mathbf{V}^0(\rho, \lambda)$; we iterate to (some measure of) convergence.

[1] This model is derived from a theory. A more general empirical model would give a better fit for the data, but would not be useful for the theory.

1. Use $\mathbf{V}^k$ and GLS to estimate the intercepts b_{0i} and the coefficients of the common time trend $h(t)$. Denote the GLS residuals of this regression as $\mathbf{e}^k$ with element $e^k_{i,t}$, and the estimate of σ^2 as

$$\left(\hat{\sigma}^2\right)^{k+1} = \frac{\mathbf{e}^{k\prime}\left(\mathbf{V}^k\right)^{-1}\mathbf{e}^k}{nT-(n+q)-2}. \tag{3}$$

   There are nT observations, n region-specific constants, q parameters in the common time trend, and 2 additional estimated parameters, λ and ρ. These account for the degrees of freedom correction in the denominator of the right side of equation 3.

2. The moment condition is $\mathbf{E}(\mathbf{ee}') - \sigma^2\mathbf{V}(\rho,\lambda) = \mathbf{0}$, where $\mathbf{O}$ is the $nT \times nT$ matrix of 0s. Replacing σ^2 and $(\mathbf{ee}')$ with the sample moments $\left(\hat{\sigma}^2\right)^{k+1}$ and $\left(\mathbf{e}^k\mathbf{e}^{k\prime}\right)$, write the moment condition as

$$\mathbf{B}^k \equiv \left[\frac{\left(\mathbf{e}^k\mathbf{e}^{k\prime}\right)}{\left(\hat{\sigma}^2\right)^{k+1}} - \mathbf{V}(\rho,\lambda)\right] = \mathbf{0}, \tag{4}$$

   We want to choose the two remaining free parameters, λ and ρ, to make the matrix of differences, $\mathbf{B}^k$, close to $\mathbf{O}$. There are many matrix norms, but the Frobenious norm has been used in this context, and is the natural analog to least squares.[2] With this norm, and denoting $B^k_{m,p}$ as the (m,p) element of $\mathbf{B}^k$, the optimization problem is

$$\min_{\lambda \geq 0 \text{ and } -1<\rho<1} \sum \left[B^k_{m,p} - v_{i,j.s}(\rho,\lambda)\right]^2 \tag{5}$$

   where the sum is over the $\frac{(nT)^2+nT}{2}$ elements of the upper triangle of the matrix $\mathbf{B}^k$. The solutions to this minimization problem are the updated estimates of λ^{k+1}, ρ^{k+1}, and thus produce the updated $\mathbf{V}^{k+1}$. Repeat step 1 with this update and iterate to convergence.

Questions:

- Does this procedure seems sensible? In particular, is the minimization problem 5 a reasonable criterion?

[2]See Cui , X., Li , C., Zhao , J., Zeng , L., Zhang , D., & Pan, J. (2016). Covariance Structure Regularization via Frobenius-Norm Discrepancy. Linear Algebra and its Applications, 510, 124-145. https://doi.org/10.1016/j.laa.2016.08.013

- What can I say about the statistical properties of these estimates (unbiasedness, efficiency, consistency?)

Comments:

- $v_{i,j.s}(\rho, \lambda)$ is linear in λ, so the first order condition to problem 5, with respect to λ, is linear in λ and the maximand is convex in λ. (Inspection of the formula for $v_{i,j.s}(\rho, \lambda)$ in Section 5 verifies these claims.) Because feasibility requires $\lambda \geq 0$, the value of λ that solves problem 5 is the maximum of 0 and the (unique) solution to the first order condition.

- $v_{i,j.s}(\rho, \lambda)$ is nonlinear in ρ, so the easiest approach may be to put a grid over the relevant range of ρ, find $\lambda(\rho)$ that solves problem 5 and then search over the grid to find the minimizing ρ.

- A GLS regression using the time series of the average emissions per period returns an estimate $\rho \approx 0.85$, so the grid over ρ should be dense in the neigborhood of this point. This regression also suggests that a reasonable starting guess is $\mathbf{V}^0(\rho = 0.85, \lambda = 0)$.

3.2 Second method of moments estimator

This method seems less plausible, but not obviously wrong. In the first step, use V^k to estimate b_{0i} and the coefficients of the common time trend $h(t)$ (but not σ^2). Denote the GLS residuals of this regression as $\mathbf{e}^k$ with element $e_{i,t}^k$. Replace the moment condition 4 with

$$\tilde{\mathbf{B}}^k \equiv \left[\left(\mathbf{e}^k \mathbf{e}^{k\prime}\right) - \sigma^2 \mathbf{V}(\rho, \lambda)\right] = \mathbf{0}.$$

Again using the Frobenius norm, write the optimization problem as

$$\min_{\sigma^2 > 0,\ \lambda \geq 0 \text{ and } -1 < \rho < 1} \sum \left[\tilde{B}_{m,p}^k - \sigma^2 v_{i,j.s}(\rho, \lambda)\right]^2.$$

The first order condition for σ^2 returns the estimator

$$\left(\sigma^2\right)^k = \frac{\sum \left[\tilde{B}_{m,p}^k v_{i,j.s}(\rho, \lambda)\right]}{\sum \left[v_{i,j.s}(\rho, \lambda)\right]^2}.$$

For $\mathbf{V}$ equal to the identity matrix this expression equals the usual SSR divided by the number of observations, but for any other covariance matrix it is unfamiliar.

4 Failed estimation strategies

I have tried two estimation strategies that return contradictory results, the first implying that $\sigma_\alpha^2 = 0$ and the second implying that $\sigma_\mu^2 = 0$. I don't believe either of those conclusions. I would like to understand why the two methods "fail".

4.1 First failure

For $\rho = 0$, I have the random effects model.[3] Using that model (only) as an analogy, I estimated the actual model in two steps.

In the first step I average time t emissions over regions ($\bar{y}_t = \frac{1}{n}\sum_i y_{it}$) and use the assumption that the regressor is a common time trend to write average emissions as

$$\bar{y}_t = B_0 + h(t) + \nu_t$$
$$\text{with}$$
$$\nu_t = \rho\nu_{t-1} + \eta_t \text{ with } \eta_t \equiv \alpha_t + \theta_t = \alpha_t + \frac{\sum_i \mu_{it}}{n},\ \sigma_\eta^2 \equiv \sigma_\alpha^2 + \frac{\sigma_\mu^2}{n} \tag{6}$$
$$\mathbf{E}\nu_t\nu_{t-s} = \sigma_\eta^2 \frac{\rho^s}{1-\rho^2}$$

I can estimate equation 6 using GLS to recover ρ and σ_η^2. For this estimation I can either use the original data and in the process recover $B_0 + h(t)$, or I can detrend the data first (recovering $B_0 + h(t)$) and then use the detrended data to recover ρ and σ_η^2. (Is one method preferred over the other?)

In the second step, I subtract average emissions (over regions) in a period from the regional emissions in that period and then drop one region to avoid a singular covariance matrix. The system is

$$y_{it} - \bar{y}_t = b_{0i} - B_0 + \left(\mu_{it} - \frac{\sum_i \mu_{it}}{n}\right) \tag{7}$$

for $i = 1, 2..n-1$ and $t = 1, 2...T$. Here there is no serial correlation, but the shocks within a period are correlated over regions. I can estimate this system

[3] For the model with $\rho = 0$ to have the standard appearance of a random effects model (at least as I have seen it presented) we need to switch the i, t indices, where i identifies a region and t identifies a time period. In my setting, each time period gets hit by a shock that is constant over regions. In the standard formulation, each region gets hit by a shock that is constant over time.

using GLS, but I also have a closed form expression for the two objects of interest, $\widehat{b_{0i}-B_0}$ and $\widehat{\frac{\sigma_\mu^2}{n}}$.

My estimate of $\frac{\sigma_\mu^2}{n}$ from the second step is greater than the estimate of $\sigma_\alpha^2+\frac{\sigma_\mu^2}{n}$ from the first step, implying the infeasible result that $\sigma_\alpha^2<0$. According to Greene's discussion of the random effects model, this type of issue can arise even when $\rho=0$. One interpretation is that the data is telling me that $\sigma_\alpha^2\approx 0$. Results from the second (also failed) attempt to estimate the model make me doubt that conclusion.

4.2 Second failure

For my second attempt, I assumed normality and used maximum likelihood. To this end, I stack the panel data into the system

$$\mathbf{y}=\mathbf{X}\beta+\mathbf{v} \text{ with } \mathbf{E}\left(\mathbf{v}\mathbf{v}'\right)=\sigma^2\mathbf{V}$$

I can write the covariance matrix $\mathbf{V}$ in closed form, $\mathbf{V}=\mathbf{V}(\rho,\lambda)$, with $\lambda\equiv\frac{\sigma_\mu^2}{\sigma_\alpha^2}$

With $N\equiv nT$ observations, $\mathbf{v}=y-\mathbf{X}\beta$, and $\mathbf{E}\mathbf{v}\mathbf{v}'=\sigma^2\mathbf{V}$, the log likelihood function is

$$\ln L=-\frac{N}{2}\left[\ln(2\pi)+\ln\sigma^2\right]-\frac{1}{2\sigma^2}\mathbf{v}'\mathbf{V}^{-1}\mathbf{v}-\frac{1}{2}\ln|\mathbf{V}|.$$

The FOC for σ^2 implies

$$\hat{\sigma}^2=\frac{\hat{\mathbf{v}}'\hat{\mathbf{V}}^{-1}\hat{\mathbf{v}}}{N}$$

with $\hat{\mathbf{v}}=y-\mathbf{X}\hat{\beta}$ and $\hat{\mathbf{V}}$ equal to the ML estimator of $\mathbf{V}$. The second term of $\ln L$ equals

$$\frac{1}{2\hat{\sigma}^2}\hat{\mathbf{v}}'\hat{\mathbf{V}}^{-1}\hat{\mathbf{v}}=\frac{N}{2\hat{\mathbf{v}}'\hat{\mathbf{V}}^{-1}\hat{\mathbf{v}}}\hat{\mathbf{v}}'\hat{\mathbf{V}}^{-1}\hat{\mathbf{v}}=\frac{N}{2}.$$

The first order condition of β implies the GLS estimator

$$\hat{\beta}=\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}.$$

Dropping the constant $-\frac{N}{2}\left[1+\ln(2\pi)\right]$, gives the concentrated log likelihood function

$$\ln L=-\frac{N}{2}\left[\ln\hat{\sigma}^2\right]-\frac{1}{2}\ln\left|\hat{\mathbf{V}}\right|. \tag{8}$$

I have closed form expressions for the derivatives of V with respect to λ and ρ. I could use those derivatives and a more sophisticated numerical routine to maximize the concentrated log likelihood function. However, because I am searching over only two parameters ρ and λ, and because I can obtain a reasonable estimate of ρ using the "first failed method", a simple grid search seems reasonable. Given an estimate of ρ, I do a grid search over $\lambda \geq 0$ and find that $\ln L$ always (i.e. for the the range of ρ that I used) decreases in λ. Therefore, the concentrated log likelihood is maximized at the estimate $\lambda = \frac{\sigma_\mu^2}{\sigma_\alpha^2} = 0$, which implies that σ_μ^2 is small relative to σ_α^2 – the opposite of the conclusion I obtained using the first failed method.

5 The covariance matrix

I need the information presented here to implement the two procedures that I describe above, but this information is not needed in order to understand the problem. It is enough to realize that I have a closed form expression for the covariance matrix.

Define

$$s = |\tau - t|$$

and

$$\begin{aligned} \iota(i,j) &= \begin{cases} 1 \text{ for } i = j \\ 0 \text{ for } i \neq j \end{cases} \\ \kappa(s) &= \begin{cases} 0 \text{ if } s = 0 \\ 1 \text{ if } s \neq 0 \end{cases} \end{aligned}$$

[I need two indicator functions to determine whether two observations correspond to the same region (in which case $\iota(i,j) = 1$ and whether they correspond to **different** time periods (in which case $\kappa(s) = 1$).]

The covariance is

$$\begin{aligned} \mathbf{E}\nu_{i,t}\nu_{j,t+s} &= \sigma_\alpha^2\rho^s\frac{1}{1-\rho^2} + \left((1-\kappa(s))\,\iota(i,j) + \kappa(s)\frac{\rho^s}{n} + \frac{\rho^{2+s}}{n}\frac{1}{1-\rho^2}\right)\sigma_\mu^2 \\ &= \sigma^2\left[\rho^s\frac{1}{1-\rho^2} + \left((1-\kappa(s))\,\iota(i,j) + \kappa(s)\frac{\rho^s}{n} + \frac{\rho^{2+s}}{n}\frac{1}{1-\rho^2}\right)\lambda\right] \\ \text{with } \sigma^2 &\equiv \sigma_\alpha^2 \text{ and } \lambda \equiv \frac{\sigma_\mu^2}{\sigma_\alpha^2}. \end{aligned}$$

Using the definition

$$\chi(\rho;i,j,s) \equiv \left((1-\kappa(s))\,\iota(i,j) + \kappa(s)\frac{\rho^s}{n} + \frac{\rho^{2+s}}{n}\frac{1}{1-\rho^2} \right), \tag{9}$$

I can write the covariance as

$$\mathbf{E}\nu_{i,t}\nu_{j,t+s} = \sigma^2\left[\rho^s\frac{1}{1-\rho^2} + \chi(\rho;i,j,s)\,\lambda\right] \tag{10}$$

Note that for $\kappa(s)=1$, i.e. for $s\neq 0$, we have $\chi(\rho;i,j,s) = \frac{\rho^s}{n} + \frac{\rho^{2+s}}{n}\frac{1}{1-\rho^2}$, i.e. χ is independent of i,j. This fact means that many elements of the covariance matrix are repeated. For example, with $n=2$ and $t=3$ (with the the first two observations corresponding to $i=1$ and $i=2$ for $t=1$, the second two observations corresponding to $i=1$ and $i=2$ for $t=2$, and so on) the upper triangle of the covariance matrix has the form

$$\mathbf{V} = \begin{bmatrix} A & B & C & C & E & E \\ & A & C & C & E & E \\ & & A & B & C & C \\ & & & A & C & C \\ & & & & A & B \\ & & & & & A \end{bmatrix}$$

with

$$\begin{aligned} A &= \tfrac{1}{1-\rho^2} + \left(1 + \tfrac{\rho^2}{n(1-\rho^2)}\right)\lambda \\ B &= \tfrac{1}{1-\rho^2} + \left(\tfrac{\rho^2}{n(1-\rho^2)}\right)\lambda \\ C &= \tfrac{\rho}{1-\rho^2} + \left(\tfrac{\rho}{n} + \tfrac{\rho^3}{n(1-\rho^2)}\right)\lambda \\ E &= \tfrac{\rho^2}{1-\rho^2} + \left(\tfrac{\rho^2}{n} + \tfrac{\rho^4}{n(1-\rho^2)}\right)\lambda = \rho C \end{aligned}$$

(Here $n=2$.) Using $E=\rho C$ I can write $\mathbf{V}$ more concisely as

$$\mathbf{V} = \begin{bmatrix} A & B & C & C & \rho C & \rho C \\ & A & C & C & \rho C & \rho C \\ & & A & B & C & C \\ & & & A & C & C \\ & & & & A & B \\ & & & & & A \end{bmatrix}.$$