Notes on estimation of the emissions model (5/6/2021)

1 Suggested steps

Aaron, please begin by making sure that you understand the model, described in Section 2. I have made two unsuccessful attempts (with the help of Andy and Sunny) to estimate this model. I would like you to reproduce their results (to confirm that the lack of success was not due to a coding error) and also to try a third method (or think of an alternative that I have not thought of).

I am not sure of the optimal sequence – whether you should reproduce Andy and Sunny's results before or after trying the third approach. Take a look at the alternatives and then we can talk.

Andy used a very simple approach, discussed in Appendix C.6 of the draft of the paper. This approach estimates two variances using separate regressions. Because the approaches are undertaken separately, there is no way to guarantee that the estimates taken together are sensible. Appendix C.6 explains the approach, and the problem. I highlighted parts of this section. I would like you to repeat this estimation to make sure that Andy got it right. I'm sending you a zipped file with Andy's work.

My second approach (implemented by Sunny) estimates the parameters of interest jointly using MLE. Section 7 describes the MLE approach and Section 8 explains what went wrong. The MLE approach (at least as I wrote it down and as Sunny implemented it) always wants to drive a ratio of variances (called λ below) to zero, implying that one variance is zero.

I'm very puzzled by this result. I don't think that this variance is zero in the real world. Also, the model (I believe) collapses to a standard "random effects" model (see e.g. Greene 14.4) if I set the autocorelation parameter $\rho = 0$ (therby producing a static model). It seems that the difficulty (the estimate of zero variance) that I encountered would also arise in that setting (the random effects setting). I'm totally mystified. Therefore, I would like you to reproduce Sunny's results – even though I have high confidence that he did everything correctly. I'm sending you a zipped file with all of Sunny's work, including the data.

Section 9 provides a glimmer of a third approach. I recently found some literature that I was previously unaware of, two papers by Anderson and by Joreskog. These papers are in the file I am sending you. I'm hoping that the methods that these papers outline will be successful. Or at least they will fail for a new reason!

2 Summary of model

Emissions in region i at time t equal

$$\begin{aligned} e_{it} &= \frac{1}{b}\left(b_{0i}+\beta x_{it}\right)+\frac{1}{b}\left(\rho\nu_{t-1}+\alpha_t+\mu_{it}\right) && (1)\\ \text{with } \nu_{it} &= \rho\nu_{t-1}+\alpha_t+\mu_{it} && (2)\end{aligned}$$

and

$$\alpha_t \sim iid\left(0,\sigma_\alpha^2\right), \mu_{i,t} \sim iid\left(0,\sigma_\mu^2\right), \mathbf{E}\left(\alpha_t\mu_{i,\tau}\right)=0\forall t,\tau. \quad (3)$$

The aggregate shock is

$$\begin{aligned} \nu_t &\equiv \frac{\sum_i \nu_{it}}{n} = \rho\nu_{t-1}+\alpha_t+\theta_t \text{ with } \theta_t \equiv \frac{\sum_i \mu_{it}}{n} && (4)\\ \nu_t &= \rho\nu_{t-1}+\eta_t \text{ with } \eta_t \equiv \alpha_t+\theta_t && (5)\end{aligned}$$

The "scaling parameter" b cannot be estimated. You can mentally set this parameter to 1. However, I have to retain it in order to eventually connect the estimation results with the theory. Eventually it might be worth thinking about what regressors should be included in the term βx_{it} in equation 1, but for the time being (mostly in the interest of simplicity) I want to incude only a common quadratic time trend, so $\beta x_{it} = a_1 t + a_2 t^2$.

The coefficients to be estimated in the regression are b_{0i} and a_1, a_2. These are easy to estimate. The AR(1) coefficient ρ is also easy to estimate. The work involves obtaining an estimate of σ_μ^2 and σ_α^2.

I'm unclear on the relative benefits of detrending the data before you use it for estimation, or detrending during the estimation process. Seems to me that the second approach is more efficient – except when the two approaches are equivalent. The first approach seems to make estimation of $\rho, \sigma_\mu^2, \sigma_\alpha^2$ easier, and I would be fine with beginning and maybe ending with this approach. (If you have counterarguments, happy to hear them.) With this approach, estimate equation 1 using OLS to obtain estimates of b_{0i} and a_1, a_2 (all divided by the scaling parameter b). Call $\hat{e}_{it}$ the resulting OLS estimate of e_{it} and now use the transformed data with the model $\tilde{e}_{it} = e_{it} - \hat{e}_{it} = \frac{1}{b}\left(\rho\nu_{t-1}+\alpha_t+\mu_{it}\right)$, where the moments of $\left(\rho\nu_{t-1}+\alpha_t+\mu_{it}\right)$ are given above.

3 The covariance matrix

Defining

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

and

$$\iota(i,j) = \begin{cases} 1 \text{ for } i = j \\ 0 \text{ for } i \neq j \end{cases} \tag{6}$$

$$\kappa(s) = \begin{cases} 0 \text{ if } s = 0 \\ 1 \text{ if } s \neq 0 \end{cases} \tag{7}$$

[Note the presence of the new indicator function. I need both a way of indicating 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 $\kappa(s) = 1$).]

I can write (after many hours of calculation) the covariance as

$$\mathbf{E}\nu_{i,t}\nu_{j,t+s} = \sigma_\alpha^2 \rho^s \frac{1}{1-\rho^2} + \left((1-\kappa(s))\,\iota(i,j) + \kappa(s)\frac{\rho^s}{n} + \frac{\rho^{2+s}}{n}\frac{1}{1-\rho^2}\right)\sigma_\mu^2 \tag{8}$$

Dividing by b^2 I have

$$\frac{\mathbf{E}\nu_{i,t}\nu_{j,t+s}}{b^2} = \sigma^2\left[\rho^s\frac{1}{1-\rho^2} + \left((1-\kappa(s))\,\iota(i,j) + \kappa(s)\frac{\rho^s}{n} + \frac{\rho^{2+s}}{n}\frac{1}{1-\rho^2}\right)\lambda\right] \tag{9}$$
$$\text{with } \sigma^2 \equiv \frac{\sigma_\alpha^2}{b^2} \text{ and } \lambda \equiv \frac{\sigma_\mu^2}{\sigma_\alpha^2}.$$

[Note that the formula for the covariance has changed.]

I denote the covariance of $\frac{\nu_{i,t}}{b}$ as $\sigma^2\mathbf{V}$. Note that the matrix **V does not include the scalar** σ^2. I can estimate ρ, σ^2 and λ, but not the scaling factor b.

For future use, note that

$$\frac{d\left(\rho^s\frac{1}{1-\rho^2} + \left(1-\kappa\iota+\kappa\frac{\rho^s}{n} + \frac{\rho^{2+s}}{n}\frac{1}{1-\rho^2}\right)\lambda\right)}{d\rho} =$$

$$\frac{1}{n}\frac{\rho^{s-1}}{(\rho^2-1)^2}\left(s\kappa + 2\rho^2 + s\rho^2 - s\rho^4 - 2s\kappa\rho^2 + s\kappa\rho^4\right)\lambda + \frac{1}{n}\frac{\rho^{s-1}}{(\rho^2-1)^2}\left(2n\rho^2 + ns - ns\rho^2\right) \tag{10}$$

and

$$\frac{d\left(\frac{d\left(\rho^s\frac{1}{1-\rho^2}+\left(1-\kappa\iota+\kappa\frac{\rho^s}{n}+\frac{\rho^{2+s}}{n}\frac{1}{1-\rho^2}\right)\lambda\right)}{d\rho}\right)}{d\lambda} \tag{11}$$

$$= \left(1-\kappa x+\kappa\frac{\rho^s}{n}+\frac{\rho^{2+s}}{n}\frac{1}{1-\rho^2}\right)$$

Remember the definitions of ι and κ in equations 6 and 7.

4 Write the panel as a system of equations

Define

$$\underset{4x1}{\mathbf{e}_t} = \begin{pmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \\ e_{4t} \end{pmatrix} \text{ and } \mathbf{v}_t = \begin{pmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ v_{4t} \end{pmatrix} \tag{12}$$

I am stacking the system in such a way that, for example, the first four elements equal emissions of the four regions in the first period.

$$\underset{6x6}{\mathbf{x}_t} = \begin{bmatrix} 1 & 0 & 0 & 0 & t & t^2 \\ 0 & 1 & 0 & 0 & t & t^2 \\ 0 & 0 & 1 & 0 & t & t^2 \\ 0 & 0 & 0 & 1 & t & t^2 \\ 0 & 0 & 0 & 0 & t & t^2 \\ 0 & 0 & 0 & 0 & t & t^2 \end{bmatrix} \text{ and } \underset{4Tx1}{\mathbf{e}} = \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \\ \vdots \\ \vdots \\ \mathbf{e}_T \end{pmatrix} \text{ and } \underset{4Tx6}{\mathbf{X}} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \vdots \\ \vdots \\ \mathbf{x}_T \end{pmatrix} \tag{13}$$

$$\text{and } \underset{6x1}{\beta} = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \phi_1 \\ \phi_2 \end{pmatrix} \text{ and } \underset{4Tx1}{\mathbf{v}} = \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \\ \vdots \\ \vdots \\ \mathbf{v}_T \end{pmatrix}, \text{ and } \mathbf{E}\left(\mathbf{v}\mathbf{v}'\right) = \sigma^2\mathbf{V} \tag{14}$$

(I defined $\sigma^2 = \frac{\sigma_\alpha^2}{b^2}$ above.) With this notation we can write the stacked system as

$$\mathbf{e} = \mathbf{X}\beta + \mathbf{v} \text{ with } \mathbf{E}\left(\mathbf{v}\mathbf{v}'\right) = \sigma^2\mathbf{V} \tag{15}$$

The first four elements of β equal the region-specific constants (the δ's); the next two elements equal the coefficients of the linear and quadratic time trend.

5 Writing V in a simple form

The $nT \times nT$ matrix (where n equals the number of regions – 4 in our application) has a particular block structure, with each block a $n \times n$ matrix. Write this matrix as

$$\mathbf{V} = \frac{1}{\sigma^2}\,\mathbf{E}\begin{bmatrix} \mathbf{v}_1\mathbf{v}_1' & \mathbf{v}_1\mathbf{v}_2' & \mathbf{v}_1\mathbf{v}_3' & \cdots\cdots & \cdots\cdots & \mathbf{v}_1\mathbf{v}_T' \\ & \mathbf{v}_2\mathbf{v}_2' & \mathbf{v}_2\mathbf{v}_3' & \cdots\cdots & \cdots\cdots & \mathbf{v}_2\mathbf{v}_T' \\ & & \mathbf{v}_3\mathbf{v}_3' & \cdots\cdots & \cdots\cdots & \mathbf{v}_3\mathbf{v}_T' \\ & & & \vdots & \vdots & \vdots \\ & & & & \vdots & \vdots \\ & & & & & \mathbf{v}_T\mathbf{v}_T' \end{bmatrix} \tag{16}$$

(The matrix is symmetric so I am showing only the diagonal and the upper blocks.)

Define J as the $n \times n$ matrix consisting of all 1's (in every element, both diagonal and off-diagonal) and define I as the $n \times n$ identity matrix. With these definitions, the $\frac{1}{\sigma^2}\mathbf{E}\mathbf{v}_t\mathbf{v}_{t+s}$ block has the following form (using equation 9).

$$\text{for } s = 0$$

$$\tfrac{1}{\sigma^2}\mathbf{E}\mathbf{v}_t\mathbf{v}_{t+s} = \left(\tfrac{1}{1-\rho^2} + \tfrac{\rho^2}{n}\tfrac{1}{1-\rho^2}\lambda\right) J + \lambda I \tag{17}$$

$$\text{For } s > 0$$

$$\tfrac{1}{\sigma^2}\left(\rho^s \tfrac{1}{1-\rho^2} + \left(\tfrac{\rho^s}{n} + \tfrac{\rho^{2+s}}{n}\tfrac{1}{1-\rho^2}\right)\lambda\right) J \tag{18}$$

As an additional consistency check I will write out the matrix for $n = 2$ and $T = 3$. Here the matrix is

$$K =$$

$$\begin{bmatrix} \frac{1}{1-\rho^2} + \left(1 + \frac{1}{n}\frac{\rho^2}{1-\rho^2}\right)\lambda & \frac{1}{1-\rho^2} + \left(\frac{1}{n}\frac{\rho^2}{1-\rho^2}\right)\lambda & \rho^1\frac{1}{1-\rho^2} + \left(\frac{\rho^1}{n} + \rho^{2+1}\frac{1}{n}\frac{1}{1-\rho^2}\right)\lambda & \rho^1 \cdots \\ \frac{1}{1-\rho^2} + \left(\frac{1}{n}\frac{\rho^2}{1-\rho^2}\right)\lambda & \frac{1}{1-\rho^2} + \left(1 + \frac{1}{n}\frac{\rho^2}{1-\rho^2}\right)\lambda & \rho^1\frac{1}{1-\rho^2} + \left(\frac{\rho^1}{n} + \rho^{2+1}\frac{1}{n}\frac{1}{1-\rho^2}\right)\lambda & \rho^1 \cdots \\ \rho^1\frac{1}{1-\rho^2} + \left(\frac{\rho^1}{n} + \rho^{2+1}\frac{1}{n}\frac{1}{1-\rho^2}\right)\lambda & \rho^1\frac{1}{1-\rho^2} + \left(\frac{\rho^1}{n} + \rho^{2+1}\frac{1}{n}\frac{1}{1-\rho^2}\right)\lambda & \frac{1}{1-\rho^2} + \left(1 + \frac{1}{n}\frac{\rho^2}{1-\rho^2}\right)\lambda & \cdots \\ \rho^1\frac{1}{1-\rho^2} + \left(\frac{\rho^1}{n} + \rho^{2+1}\frac{1}{n}\frac{1}{1-\rho^2}\right)\lambda & \rho^1\frac{1}{1-\rho^2} + \left(\frac{\rho^1}{n} + \rho^{2+1}\frac{1}{n}\frac{1}{1-\rho^2}\right)\lambda & \frac{1}{1-\rho^2} + \left(\frac{1}{n}\frac{\rho^2}{1-\rho^2}\right)\lambda & \cdots \\ \rho^2\frac{1}{1-\rho^2} + \left(\frac{\rho^2}{n} + \rho^{2+2}\frac{1}{n}\frac{1}{1-\rho^2}\right)\lambda & \rho^2\frac{1}{1-\rho^2} + \left(\frac{\rho^2}{n} + \rho^{2+2}\frac{1}{n}\frac{1}{1-\rho^2}\right)\lambda & \rho^1\frac{1}{1-\rho^2} + \left(\frac{\rho^1}{n} + \rho^{2+1}\frac{1}{n}\frac{1}{1-\rho^2}\right)\lambda & \rho^1 \cdots \\ \rho^2\frac{1}{1-\rho^2} + \left(\frac{\rho^2}{n} + \rho^{2+2}\frac{1}{n}\frac{1}{1-\rho^2}\right)\lambda & \rho^2\frac{1}{1-\rho^2} + \left(\frac{\rho^2}{n} + \rho^{2+2}\frac{1}{n}\frac{1}{1-\rho^2}\right)\lambda & \rho^1\frac{1}{1-\rho^2} + \left(\frac{\rho^1}{n} + \rho^{2+1}\frac{1}{n}\frac{1}{1-\rho^2}\right)\lambda & \rho^1 \cdots \end{bmatrix} \tag{19}$$

Because the matrix does not fit on a page, I write it for $\rho = 0.9$ and $n = 2$

$$\begin{bmatrix} 3.131578947\lambda + 5.263157895 & 2.131578947\lambda + 5.263157895 & 2.368421053\lambda + 4.73684210\cdots \\ 2.131578947\lambda + 5.263157895 & 3.131578947\lambda + 5.263157895 & 2.368421053\lambda + 4.73684210\cdots \\ 2.368421053\lambda + 4.736842105 & 2.368421053\lambda + 4.736842105 & 3.131578947\lambda + 5.26315789\cdots \\ 2.368421053\lambda + 4.736842105 & 2.368421053\lambda + 4.736842105 & 2.131578947\lambda + 5.26315789\cdots \\ 2.131578947\lambda + 4.263157895 & 2.131578947\lambda + 4.263157895 & 2.368421053\lambda + 4.73684210\cdots \\ 2.131578947\lambda + 4.263157895 & 2.131578947\lambda + 4.263157895 & 2.368421053\lambda + 4.73684210\cdots \end{bmatrix} \tag{20}$$

The determinant: $d = 5.263157866\lambda^6 + 31.57894727\lambda^5 + 63.1578947\lambda^4 + 42.10526324\lambda^3$

The matrix K is V evaluated at $\rho = 0.9$, $n = 2$ and $T = 3$

To make this even easier to read, I write it for $\lambda = 2.4$ (an arbitrary number). Now the matrix is

$$\begin{bmatrix} 12.77894737 & 10.37894737 & 10.42105263 & 10.42105263 & 9.378947368 & 9.378947368 \\ 10.37894737 & 12.77894737 & 10.42105263 & 10.42105263 & 9.378947368 & 9.378947368 \\ 10.42105263 & 10.42105263 & 12.77894737 & 10.37894737 & 10.42105263 & 10.42105263 \\ 10.42105263 & 10.42105263 & 10.37894737 & 12.77894737 & 10.42105263 & 10.42105263 \\ 9.378947368 & 9.378947368 & 10.42105263 & 10.42105263 & 12.77894737 & 10.37894737 \\ 9.378947368 & 9.378947368 & 10.42105263 & 10.42105263 & 10.37894737 & 12.77894737 \end{bmatrix}$$

Check to confirm that the block are the same as shown in equations 17 and 18.

For $s=0$ I have the block

$$\left(\frac{1}{1-\rho^2}+\frac{\rho^2}{n}\frac{1}{1-\rho^2}\lambda\right)\left[\begin{array}{cc}1 & 1\\ 1 & 1\end{array}\right]+\lambda\left[\begin{array}{cc}1 & 0\\ 0 & 1\end{array}\right]=$$

$$\left[\begin{array}{cc}3.131\,578\,947\lambda+5.263\,157\,895 & 2.131\,578\,947\lambda+5.263\,157\,895\\ 2.131\,578\,947\lambda+5.263\,157\,895 & 3.131\,578\,947\lambda+5.263\,157\,895\end{array}\right] \tag{21}$$

For $s=1$ I have the block

$$\left(\rho^s\frac{1}{1-\rho^2}+\left(\frac{\rho^s}{n}+\frac{\rho^{2+s}}{n}\frac{1}{1-\rho^2}\right)\lambda\right)\left[\begin{array}{cc}1 & 1\\ 1 & 1\end{array}\right]=$$

$$\left[\begin{array}{cc}2.368\,421\,053\lambda+4.736\,842\,105 & 2.368\,421\,053\lambda+4.736\,842\,105\\ 2.368\,421\,053\lambda+4.736\,842\,105 & 2.368\,421\,053\lambda+4.736\,842\,105\end{array}\right] \tag{22}$$

For $s=2$ I have the block

$$\left(\rho^s\frac{1}{1-\rho^2}+\left(\frac{\rho^s}{n}+\frac{\rho^{2+s}}{n}\frac{1}{1-\rho^2}\right)\lambda\right)\left[\begin{array}{cc}1 & 1\\ 1 & 1\end{array}\right]=$$

$$\left[\begin{array}{cc}2.131\,578\,947\lambda+4.263\,157\,895 & 2.131\,578\,947\lambda+4.263\,157\,895\\ 2.131\,578\,947\lambda+4.263\,157\,895 & 2.131\,578\,947\lambda+4.263\,157\,895\end{array}\right] \tag{23}$$

6 The old procedure for constructing V

Before I realized that **V** has the simple form given in the previous section, I invented more complicated notation to construct it. I include this material but I think that it is superseded by the previous section. In any case it is useful as a consistency check for the method in the previous section. However, at first reading this section can be skipped.)

I want a procedure for associating an arbitrary element of the vector **v** with a particular time period and region. (It seems to me that we need this information for programming....) Define the floor function $\lfloor y \rfloor$ as the largest integer no greater than y. For example, $\lfloor 7.2 \rfloor = 7$ and $\lfloor 7.0 \rfloor = 7$, and $\lfloor 0.25 \rfloor = 0$. Consider v_m, defined as the m'th element of $\mathbf{v}$. This element corresponds to time period $t = \left\lfloor \frac{m-1}{4} \right\rfloor + 1$ and region $j = m - 4(t-1)$. (Check!)

We can also invert this relation. For example, observation t, j corresponds to $m = (t-1)\,4 + j$. With this information we can translate the elements defined by equation 9 into elements of the matrix $\mathbf{V}$.

The (m, p) element, $\mathbf{V}_{m,p}$ is

$$\mathbf{E}v_m v_p = \mathbf{E}v_{m-4(t-1),t}v_{p-4(\tau-1),\tau} \text{ with } t = \left\lfloor \frac{m-1}{4} \right\rfloor + 1 \text{ and } \tau = \left\lfloor \frac{p-1}{4} \right\rfloor + 1 \tag{24}$$

(I see that this notation might be ambiguous. By $\mathbf{E}v_m v_p$ I mean the expectation of the product of the m'th and the p'th element of the $4T \times 1$ vector $\mathbf{v}$. By $\mathbf{E}v_{m-4(t-1),t}v_{p-4(\tau-1),\tau}$ I mean the expectation of the product of the errors associated with time period t and region $m - 4(t-1)$, and time period τ and region $p - 4(\tau - 1)$.)

Above I defined the absolute value of the difference between two time indices as s. Using the expressions for any two arbitrary elements of the vector of stacked errors, $\mathbf{v}$, m and p, and equation 24, I have

$$\tau - t = \left\lfloor \frac{p-1}{4} \right\rfloor + 1 - \left(\left\lfloor \frac{m-1}{4} \right\rfloor + 1 \right) = \left\lfloor \frac{p-1}{4} \right\rfloor - \left\lfloor \frac{m-1}{4} \right\rfloor. \tag{25}$$

Therefore,

$$s \equiv |\tau - t| = \left| \left\lfloor \frac{p-1}{4} \right\rfloor - \left\lfloor \frac{m-1}{4} \right\rfloor \right| \tag{26}$$

Do a consistency check. Suppose that I want to identify the location in the $\mathbf{V}$ matrix of the covariance between the shock to region $j = 2$ in period $t = 6$ and region $k = 4$ in period $\tau = 12$. This element is in row $m = (t-1)\,4 + j = (6-1)\,4 + 2 = 22$ and in column $p = (\tau - 1)\,4 + k = (12-1)\,4 + 4 = 48$. Not make sure that the inversion works. Use $t = \left\lfloor \frac{m-1}{4} \right\rfloor + 1 = \left\lfloor \frac{22-1}{4} \right\rfloor + 1 = 6$ and $\tau = \left\lfloor \frac{p-1}{4} \right\rfloor + 1 = \left\lfloor \frac{48-1}{4} \right\rfloor + 1 = 12$. OK, I am picking up the right time indices. For agent i use $j = m - 4(t-1) = 22 - 4(6-1) = 2$ and for agent k use $k = p - 4(\tau - 1) = 48 - 4(12-1) = 4$. OK, I am picking up the right agent indices.

7 The Maximum Liklihood estimation procedure

Define $\Gamma = \mathbf{V}^{-1}$. Greene calls the covariance matrix Ω. I call this matrix $\mathbf{V}$. Greene calls the vector of LHS variables y. I call this vector $\mathbf{e}$. Greene assumes that there are n observations. For my problem there are $4T$ observations. Green calls the vector of residuals $\boldsymbol{\epsilon}$. I refer to these residuals as $\tilde{\mathbf{v}} = \mathbf{e} - \mathbf{X}\tilde{\beta}$. Of couse I am using my notation, so that these notes are self-contained.

Greene pg 471 gives the estimation equations for β and σ as the solutions to the first order conditions (from minimizing the likelihood function, assuming normality) as

$$\mathbf{X}'\Gamma\left(\mathbf{e} - \mathbf{X}\boldsymbol{\beta}\right) = 0 \tag{27}$$

$$-\frac{4T}{2\sigma^2} + \frac{1}{2\sigma^4}\left(\mathbf{e} - \mathbf{x}\boldsymbol{\beta}\right)'\Gamma\left(\mathbf{e} - \mathbf{x}\boldsymbol{\beta}\right) = 0. \tag{28}$$

The derivative of the likelihood function with respect to Γ returns the first order condition (eq 11-31, pg 471 in Greene).

$$\frac{1}{2\sigma^2}\left(\sigma^2\mathbf{V} - \tilde{\mathbf{v}}\tilde{\mathbf{v}}'\right) = 0. \tag{29}$$

Greene notes that the covariance matrix must be restricted in some way, i.e. it must be possible to write the covariance matrix as $\mathbf{V} = \mathbf{V}(\gamma)$, where γ is a vector of parameters. For my problem, $\gamma = (\rho, \lambda)$. (Greene refers to this parameter vector as θ, but I used that symbol above for a different purpose.) Greene outlines the Oberhofer-Kmenta algorithm for achieving consistent estimators: begin with a consistent estimator of γ. Using these, solve equations 27 and 28 to obtain estimates of β and σ^2. Using these, solve equation 29 to obtain an estimate of γ. Repeat until satisfactory convergence. He gives conditions (pg 472) for this process to converge to the MLE. The only condition that I do not know how to satisfy is the requirement that the starting guess for γ is consistent. I do not know how to guarantee this. A practical solution would be to begin with different initial values of the guess for γ and show that the convergence does not depend on the initial guess.

The remaining issue is how to solve 29 to obtain an estimate of γ, conditional on estimates of β and σ^2. There may be a clever way of doing this without using derivatives. However, because we have a fairly simple expression for $\mathbf{V}$ it might be easier to use derivatives. Define L as the likelihood function.

$$L = \text{"}terms\text{"} - \frac{1}{2\sigma^2}\tilde{\mathbf{v}}'\mathbf{\Gamma}\tilde{\mathbf{v}} + \frac{1}{2}\ln|\Gamma| \tag{30}$$

where "terms" are independent of Γ. Use the rule for the derivative of the inverse of a matrix:

$$\frac{dA^{-1}}{d\rho} = -A^{-1}\frac{dA}{d\rho}A^{-1}. \tag{31}$$

Consider the FOC for ρ. I have

$$\frac{\partial\tilde{\mathbf{v}}'\mathbf{\Gamma}\tilde{\mathbf{v}}}{\partial\rho} = \frac{\partial\tilde{\mathbf{v}}'\mathbf{V}^{-1}\tilde{\mathbf{v}}}{\partial\rho} = -\tilde{\mathbf{v}}'\mathbf{V}^{-1}\frac{d\mathbf{V}}{d\rho}\mathbf{V}^{-1}\tilde{\mathbf{v}} \tag{32}$$

I also have the rule (see "cookbook" eqn 10)

$$\begin{aligned} \partial\left(\ln|\Gamma|\right) &= Tr\left(\Gamma^{-1}\partial\Gamma\right) \Rightarrow \frac{\partial\left(\ln|\Gamma|\right)}{\partial\rho} = Tr\left(\Gamma^{-1}\frac{\partial\Gamma}{\partial\rho}\right) & (33) \\ &= -Tr\left(\mathbf{V}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\rho}\mathbf{V}^{-1}\right) = -Tr\left(\frac{\partial\mathbf{V}}{\partial\rho}\mathbf{V}^{-1}\right) & (34) \end{aligned}$$

Using these two equations I have the FOC for ρ:

$$\frac{d\ln L}{d\rho} = \frac{1}{2\sigma^2}\tilde{\mathbf{v}}'\mathbf{V}^{-1}\frac{d\mathbf{V}}{d\rho}\mathbf{V}^{-1}\tilde{\mathbf{v}} - Tr\left(\frac{d\mathbf{V}}{d\rho}\mathbf{V}^{-1}\right) = 0 \tag{35}$$

I also have the FOC for λ

$$\frac{d\ln L}{d\lambda} = \frac{1}{2\sigma^2}\tilde{\mathbf{v}}'\mathbf{V}^{-1}\frac{d\mathbf{V}}{d\lambda}\mathbf{V}^{-1}\tilde{\mathbf{v}} - Tr\left(\frac{d\mathbf{V}}{d\lambda}\mathbf{V}^{-1}\right) = 0 \tag{36}$$

Conditional on estimates of σ^2 and β, equations 35 and 35 comprise two equations in two unkowns, ρ and λ.

We can estimate the parameters using the method of steepest gradient ascent (see https://en.wikipedia.org/wiki/Gradient_descent). Denote $a = (\rho, \lambda)$, the parameters we want to estimate (along with β and σ^2). Denote $F = \ln L$, our maximand, and denote $\nabla F(a)$, the gradient. The first element of $\nabla F(a)$ is given by the left side of equation 35 and the second element is given by the left side of equation 36. (I'm using these definitions so that things line up with the Wikipedia article.) Here is the proposed algorithm.

1. We need an initial guess to begin the method of gradient ascent. $\rho_0 = 0.9$ is a reasonable starting value for ρ. (The subscript 0 denotes the starting value.) To find a reasonable starting value for λ, use a grid, e.g. $\lambda \in \{0.5, 1, 1.5, 2....20\}$. For each of these values calculate V and V^{-1}. Using these matrices calculate the estimate of β and σ^2 :

$$\hat{\beta} = \left(\mathbf{X}'V^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'V^{-1}\mathbf{e} \tag{37}$$

$$\hat{\sigma}^2 = \frac{1}{n}\left(\mathbf{e} - \mathbf{X}'\beta\right)' V^{-1}\left(\mathbf{e} - \mathbf{X}'\beta\right) \tag{38}$$

with $n = 4T$. (**Whoops!** Sloppy notation. Everywhere else in this note I use n to denote the number of regions, $n = 4$ for our problem. But here I use it to denote $4T$.) For each of these values of λ, calculate the log likelihood function

$$\ln L = -\frac{n}{2}\left[\ln\left(2\pi + \ln \sigma^2\right)\right] - \frac{1}{2\sigma^2}\tilde{\mathbf{v}}'V^{-1}\tilde{\mathbf{v}} + \frac{1}{2}\ln\left|V^{-1}\right| \tag{39}$$

with $\tilde{\mathbf{v}} = \mathbf{e} - \mathbf{X}\tilde{\beta}$. Choose the λ that maximizes the criteria; call this value λ_0. **Graph $\ln L$ over λ to see if the function is well behaved.** If it reaches a maximum at either 0.5 or 20, you need to change the initial grid to go beyond these points.

2. Using $a_0 = (\rho_0, \lambda_0)$, update the guess using $a_{n+1} = a_n + \gamma_n \nabla F(a_n)$, where $\gamma_n > 0$ is the step size at iteration n. The Wikipedia article gives the formula for γ_n, called the Barzilai-Borwein method. This looks simple to calculate.

3. Stop when $\parallel a_j - a_{j-1} \parallel$ is small. (I don't know what small means in this context; we have to experiment.)

I think that perhaps the greatest danger of a coding error arises in correctly assigning the entries of the V matrix and the derivative of this matrix. Therefore, I suggest writing the code where T (the number of periods) and n (the number of regions) are parameters. To check the code, set $T = 3$ and $n = 2$, so that V and the derivative matrices $\frac{d\mathbf{V}}{d\rho}$ and $\frac{d\mathbf{V}}{d\lambda}$ are all 6x6 matrices. You can construct this matrix by hand, and then confirm that your code returns the correct matrices.

8 Lessons from Sunny's estimation

Sunny finds that the log likelihood decreases in λ, so the log likelhood is maximized at $\lambda = 0$. This section tries to understand the effect of λ on $\ln L$, given ρ. With $N \equiv nT$ observations, $\mathbf{v} = \mathbf{e} - \mathbf{X}\beta$, and $\mathbf{E}\mathbf{v}\mathbf{v}' = \sigma^2\mathbf{V}$, the log likelihood function is

$$\begin{aligned} \ln L &= -\tfrac{N}{2}\left[\ln\left(2\pi + \ln\sigma^2\right)\right] - \tfrac{1}{2\sigma^2}\mathbf{v}'\mathbf{V}^{-1}\mathbf{v} + \tfrac{1}{2}\ln\left|\mathbf{V}^{-1}\right| \\ &= -\tfrac{N}{2}\left[\ln\left(2\pi\right) + \ln\sigma^2\right] - \tfrac{1}{2\sigma^2}\mathbf{v}'\mathbf{V}^{-1}\mathbf{v} - \tfrac{1}{2}\ln\left|\mathbf{V}\right|. \end{aligned} \tag{40}$$

The last equality uses $\ln\left|\mathbf{V}^{-1}\right| = -\ln\left|\mathbf{V}\right|$.

The FOC for σ^2 is

$$\frac{\partial \ln L}{\partial \sigma^2} = \frac{-N}{2\sigma^2} + \frac{1}{2\sigma^4}\mathbf{v}'\mathbf{V}^{-1}\mathbf{v} = \mathbf{0} \Rightarrow \hat{\sigma}^2 = \frac{\hat{\mathbf{v}}'\hat{\mathbf{V}}^{-1}\hat{\mathbf{v}}}{N} \tag{41}$$

with $\hat{\mathbf{v}} = \mathbf{e} - \mathbf{X}\hat{\beta}$ and $\hat{\mathbf{V}}$ equal to the ML estimator of $\mathbf{V}$ (i.e. the hats indicate estimated values). Using equation 41 in the second term of equation 40 gives

$$\frac{1}{2\hat{\sigma}^2}\hat{\mathbf{v}}'\hat{\mathbf{V}}^{-1}\hat{\mathbf{v}} = \frac{N}{2\hat{\mathbf{v}}'\hat{\mathbf{V}}^{-1}\hat{\mathbf{v}}}\hat{\mathbf{v}}'\hat{\mathbf{V}}^{-1}\hat{\mathbf{v}} = \frac{N}{2}. \tag{42}$$

Thus, the second term in $\ln L$ is minus $\frac{N}{2}$. Sunny's example with $\rho = 0.9$ returns this result.

The first order condition of β implies the GLS estimator

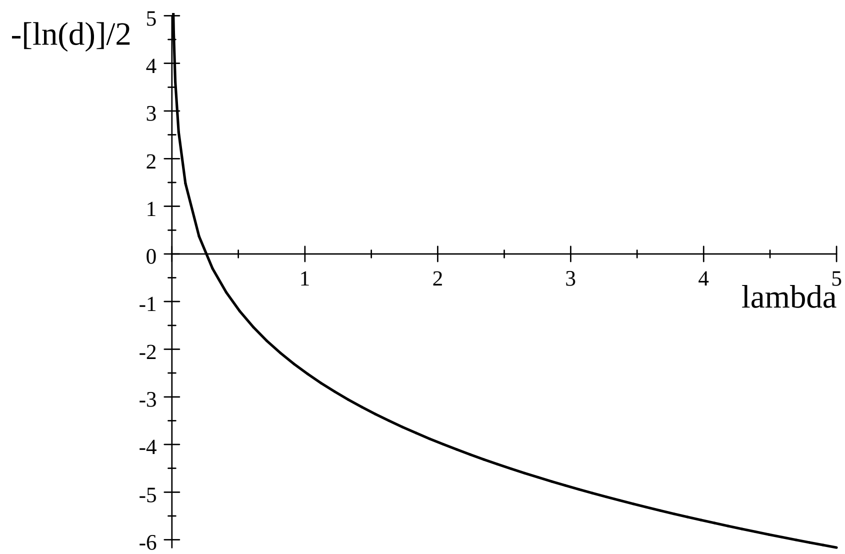
$$\hat{\beta} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{e}, \tag{43}$$

but this formula is not going to help me understand Sunny's results.

Substituting equations 41 and 42 into equation 40, and dropping the constant $-\frac{N}{2}\left[1 + \ln\left(2\pi\right)\right]$, gives

$$\ln L = -\frac{N}{2}\left[\ln\hat{\sigma}^2\right] - \frac{1}{2}\ln\left|\hat{\mathbf{V}}\right| \tag{44}$$

I know that as $\lambda \to 0$, $|\mathbf{V}| \to 0$, so as $\lambda \to 0$, $\ln|\mathbf{V}| \to -\infty$ and $-\ln|\mathbf{V}| \to \infty$. For my example above with $n = 2$ and $T = 3$ and $\rho = 0.9$, the determinant of V is $d = 5.263\,157\,866\lambda^6 + 31.578\,947\,27\lambda^5 + 63.157\,894\,7\lambda^4 + 42.105\,263\,24\lambda^3$. The last term of the liklihood function (second line) is $-\frac{1}{2}\ln d$, which decreases in λ as Figure 8 illustrates.



I don't find it surprsing that $-\frac{1}{2}\ln\left|\hat{\mathbf{V}}\right|$ decreases in λ for general n and T, so Sunny's results for the third term of the liklihood function are believable. I don't have a sense of the relation between $\left[\ln\hat{\sigma}^2\right]$ and λ, but Sunny's experiments show that this term also decreases with λ. With both $-\frac{N}{2}\left[\ln\hat{\sigma}^2\right]$ and $-\frac{1}{2}\ln\left|\hat{\mathbf{V}}\right|$ decreasing with λ, it follows that $\lambda = 0$ maximizes the log liklihood (subject to $\lambda \geq 0$).

9 The third approach

I have not written this approach up in a clear way, but the basic idea is to follow the literature that focuses on estimation of a covariance matrix. "Most econometrics" (I think) focuses on estimations of coefficients in a regression, but this smaller (I think) literature focuses on estimation of the covariance matrix. The two papers that I found are by Anderson (73) and Joreskog (78). Anderson has a textbook from the early 2000's, which presumably will cover this material, but I do not have this book. Before spending a lot of time trying to understand these two papers, it makes sense to look for a modern treatment. Conditional on an estimate of ρ, my estimation problem has the form assumed by Anderson, so his approach could be used. Joreskog's generalization permits the estimation of ρ jointly with the estimation of the variances of interest.

I emphasize that it is worthwhile finding a somewhat modern textbook treatment of this material. Ethan or Aprajit are the most likely sources of information about that material.

Ignore following material.

Extra notes.............

The basic model is

$$e_{it}=\frac{1}{b}\left(b_{0i}+h\left(t\right)+\nu_{i,t}\right), \tag{45}$$

with

$$\begin{gathered}\text{(i) RAE: }\nu_t=\rho\nu_{t-1}+\theta_t\\ \text{(ii) FE: }\nu_{i,t}=\rho\nu_{t-1}+\alpha_t+\mu_{i,t}\text{, with}\\ \text{(iii) }\alpha_t\sim iid\left(0,\sigma_\alpha^2\right),\ \mu_{i,t}\sim iid\left(0,\sigma_\mu^2\right),\ \mathbf{E}\left(\alpha_t\mu_{i,\tau}\right)=0\forall t,\tau\text{, and}\\ \text{(iv) }\theta_t=\alpha_t+\tfrac{\sum_j\mu_{jt}}{n}\text{, with }var\left(\theta_t\right)\equiv\sigma^2=\sigma_\alpha^2+\tfrac{\sigma_\mu^2}{n}.\end{gathered} \tag{46}$$

$$\begin{aligned}\nu_{i,t}=\rho\nu_{t-1}+\alpha_t+\mu_{i,t}&=\rho\nu_{t-1}+\alpha_t+\tfrac{\sum_j\mu_{jt}}{n}-\tfrac{\sum_j\mu_{jt}}{n}+\mu_{i,t}\\ &=\nu_t-\tfrac{\sum_j\mu_{jt}}{n}+\mu_{i,t}\end{aligned} \tag{47}$$

Write the regression 45 as

$$e_{it}=\frac{1}{b}\left(b_{0i}+h\left(t\right)+\rho\nu_{t-1}+\alpha_t+\mu_{i,t}\right) \tag{48}$$

$$=\frac{1}{b}\left(b_{0i}+h\left(t\right)+\rho\nu_{t-1}+\alpha_t+\frac{\sum_j\mu_{jt}}{n}-\frac{\sum_j\mu_{jt}}{n}+\mu_{i,t}\right) \tag{49}$$

$$=\frac{1}{b}\left(b_{0i}+h\left(t\right)+\nu_t+\mu_{i,t}-\frac{\sum_j\mu_{jt}}{n}\right), \tag{50}$$

Lag and multiply by ρ

$$0=-\rho e_{i,t-1}+\frac{1}{b}\rho\left(b_{0i}+h\left(t-1\right)\right)+\frac{\rho}{b}\left(\nu_{t-1}+\mu_{i,t-1}-\frac{\sum_j\mu_{j,t-1}}{n}\right) \tag{51}$$

Subtract equations 45 and 51, using $\nu_{i,t} = \rho\nu_{t-1} + \alpha_t + \mu_{i,t}$

$$\begin{aligned} e_{it} &= \tfrac{1}{b}\left(b_{0i} + h\left(t\right) + \rho\nu_{t-1} + \alpha_t + \mu_{i,t}\right) + \rho e_{i,t-1} \\ &-\tfrac{1}{b}\rho\left(b_{0i} + h\left(t-1\right)\right) - \tfrac{\rho}{b}\left(\nu_{t-1} - \tfrac{\sum_j \mu_{j,t-1}}{n} + \mu_{i,t-1}\right) = \\ \rho e_{i,t-1} &+ \tfrac{1-\rho}{b}b_{0i} + \tfrac{1}{b}\left(h\left(t\right) - \rho h\left(t-1\right)\right) + \tfrac{1}{b}\left(\alpha_t + \mu_{i,t} + \rho\left(\mu_{i,t-1} - \tfrac{\sum_j \mu_{j,t-1}}{n}\right)\right) \end{aligned} \tag{52}$$

Define

$$\epsilon_{i,t} = \left(\alpha_t + \mu_{i,t} + \rho\left(\mu_{i,t-1} - \frac{\sum_j \mu_{j,t-1}}{n}\right)\right) \tag{53}$$

$$\mathbf{E}\left(\epsilon_{i,t}\epsilon_{j,t-s}\right) = \begin{cases} \sigma_\alpha^2 + \sigma_\mu^2\left(1 + \rho^2\left(1 - \frac{1}{n}\right)\right) \text{ if } i = j \text{ and } s = 0 \\ \sigma_\alpha^2 - \rho^2\sigma_\mu^2\frac{1}{n} \text{ if } i \neq j \text{ and } s = 0 \\ \rho\sigma_\mu^2\left(1 - \frac{1}{n}\right) \text{ if } i = j \text{ and } s = 1 \\ -\sigma_\mu^2\rho\frac{1}{n} \text{ if } i \neq j \text{ and } s = 1 \\ 0 \text{ if } s > 1 \end{cases} \tag{54}$$

$$\mathbf{E}\left(\epsilon_{i,t}\epsilon_{j,t-s}\right) = \sigma_\alpha^2\begin{cases} 1 + \lambda\left(1 + \rho^2\left(1 - \frac{1}{n}\right)\right) \text{ if } i = j \text{ and } s = 0 \\ 1 - \rho^2\lambda\frac{1}{n} \text{ if } i \neq j \text{ and } s = 0 \\ \rho\lambda\left(1 - \frac{1}{n}\right) \text{ if } i = j \text{ and } s = 1 \\ -\lambda\rho\frac{1}{n} \text{ if } i \neq j \text{ and } s = 1 \\ 0 \text{ if } s > 1 \end{cases}$$

$$\Omega = \begin{bmatrix} 1 + \lambda\left(1 + \rho^2\left(1 - \frac{1}{n}\right)\right) & 1 - \rho\frac{1}{n}\lambda & 1 - \rho\frac{1}{n}\lambda & 1 - \rho\frac{1}{n}\lambda \\ 1 - \rho\frac{1}{n}\lambda & 1 + \lambda\left(1 + \rho^2\left(1 - \frac{1}{n}\right)\right) & 1 - \rho\frac{1}{n}\lambda & 1 - \rho\frac{1}{n}\lambda \\ 1 - \rho\frac{1}{n}\lambda & 1 - \rho\frac{1}{n}\lambda & 1 + \lambda\left(1 + \rho^2\left(1 - \frac{1}{n}\right)\right) & 1 - \rho\frac{1}{n}\lambda \\ 1 - \rho\frac{1}{n}\lambda & 1 - \rho\frac{1}{n}\lambda & 1 - \rho\frac{1}{n}\lambda & 1 + \left(1 + \rho^2\right)\lambda \end{bmatrix}$$

For $\lambda = 0$, Ω is singular, so for any ρ I'm still going to get a singular matrix (for the covariance matrix for the entire system). If I use ML, setting $\lambda = 0$ will probably still maximize the likelihood. It is clear why the covariance matrix is singular at $\lambda = 0$ because then the shocks in a given period are perfectly correlated. There is no way to get around that.

$$\Omega = \begin{bmatrix} a & c & c & c \\ c & a & c & c \\ c & c & a & c \\ c & c & c & a \end{bmatrix}$$

$$\Omega\text{, inverse: } \begin{bmatrix} \frac{a+2c}{a^2+2ac-3c^2} & -\frac{c}{a^2+2ac-3c^2} & -\frac{c}{a^2+2ac-3c^2} & -\frac{c}{a^2+2ac-3c^2} \\ -\frac{c}{a^2+2ac-3c^2} & \frac{a+2c}{a^2+2ac-3c^2} & -\frac{c}{a^2+2ac-3c^2} & -\frac{c}{a^2+2ac-3c^2} \\ -\frac{c}{a^2+2ac-3c^2} & -\frac{c}{a^2+2ac-3c^2} & \frac{a+2c}{a^2+2ac-3c^2} & -\frac{c}{a^2+2ac-3c^2} \\ -\frac{c}{a^2+2ac-3c^2} & -\frac{c}{a^2+2ac-3c^2} & -\frac{c}{a^2+2ac-3c^2} & \frac{a+2c}{a^2+2ac-3c^2} \end{bmatrix}$$

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Find the covariance terms

$i = j$ and $s = 0$

$$\mathbf{E}\left(\alpha_t + \mu_{i,t} + \rho\left(\mu_{i,t-1} - \frac{\sum_j \mu_{j,t-1}}{n}\right)\right)\left(\alpha_t + \mu_{i,t} + \rho\left(\mu_{i,t-1} - \frac{\sum_j \mu_{j,t-1}}{n}\right)\right) = \sigma_\alpha^2 + \sigma_\mu^2\left(1 + \rho^2\left(1 - \frac{1}{n}\right)\right)$$

$$\rho^2\mathbf{E}\left(\left(\mu_{i,t-1} - \frac{\sum_j \mu_{j,t-1}}{n}\right)\right)\left(\left(\mu_{i,t-1} - \frac{\sum_j \mu_{j,t-1}}{n}\right)\right) = \rho^2\sigma_\mu^2\left(1 - \frac{2}{n} + \frac{n}{n^2}\right) = \rho^2\sigma_\mu^2\left(1 - \frac{2}{n} + \frac{1}{n}\right) = \rho^2\sigma_\mu^2\left(1 - \frac{1}{n}\right)$$

$i = j$ and $s = 1$

$$\mathbf{E}\left(\alpha_t + \mu_{i,t} + \rho\left(\mu_{i,t-1} - \frac{\sum_j \mu_{j,t-1}}{n}\right)\right)\left(\alpha_{t-1} + \mu_{i,t-1} + \rho\left(\mu_{i,t-2} - \frac{\sum_j \mu_{j,t-2}}{n}\right)\right) = \sigma_\mu^2\rho\left(1 - \frac{1}{n}\right)$$

$i \neq j$ and $s = 0$

$$\mathbf{E}\left(\alpha_t + \mu_{i,t} + \rho\left(\mu_{i,t-1} - \frac{\sum_j \mu_{j,t-1}}{n}\right)\right)\left(\alpha_t + \mu_{j,t} + \rho\left(\mu_{j,t-1} - \frac{\sum_j \mu_{j,t-1}}{n}\right)\right) = \sigma_\alpha^2 - \rho^2\sigma_\mu^2\frac{1}{n}$$

use

$$\mathbf{E}\left(\mu_{i,t} + \rho\left(\mu_{i,t-1} - \frac{\sum_j \mu_{j,t-1}}{n}\right)\right)\left(\mu_{j,t} + \rho\left(\mu_{j,t-1} - \frac{\sum_j \mu_{j,t-1}}{n}\right)\right) = \mathbf{E}\left(\rho\left(\mu_{i,t-1} - \frac{\sum_j \mu_{j,t-1}}{n}\right)\right)\left(\rho\left(\mu\right.\right.$$

$$= \rho^2\mathbf{E}\left(\left(\mu_{i,t-1} - \frac{\sum_j \mu_{j,t-1}}{n}\right)\right)\left(\left(\mu_{j,t-1} - \frac{\sum_j \mu_{j,t-1}}{n}\right)\right) = \rho^2\sigma_\mu^2\left(-\frac{2}{n} + \frac{1}{n}\right) = -\rho^2\sigma_\mu^2\frac{1}{n}$$

$i \neq j$ and $s = 1$

$$\mathbf{E}\left(\alpha_t + \mu_{i,t} + \rho\left(\mu_{i,t-1} - \frac{\sum_j \mu_{j,t-1}}{n}\right)\right)\left(\alpha_{t-1} + \mu_{j,t-1} + \rho\left(\mu_{j,t-2} - \frac{\sum_j \mu_{j,t-2}}{n}\right)\right) = -\sigma_\mu^2\rho\frac{1}{n}$$

use

$\mathbf{E}\left(\rho\left(-\frac{\sum_j \mu_{j,t-1}}{n}\right)\right)\left(\mu_{j,t-1}\right) = -\sigma_\mu^2 \rho \frac{1}{n}$
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

xxxxxxxxxxxxxxxxxxxxxxxxx
Write it out in detail

For $s = 0$

$$\begin{bmatrix} \frac{1}{1-\rho^2} + \left(1 + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\right)\lambda & \frac{1}{1-\rho^2} + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\lambda & \frac{1}{1-\rho^2} + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\lambda & \frac{1}{1-\rho^2} + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\lambda \\ \frac{1}{1-\rho^2} + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\lambda & \frac{1}{1-\rho^2} + \left(1 + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\right)\lambda & \frac{1}{1-\rho^2} + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\lambda & \frac{1}{1-\rho^2} + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\lambda \\ \frac{1}{1-\rho^2} + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\lambda & \frac{1}{1-\rho^2} + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\lambda & \frac{1}{1-\rho^2} + \left(1 + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\right)\lambda & \frac{1}{1-\rho^2} + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\lambda \\ \frac{1}{1-\rho^2} + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\lambda & \frac{1}{1-\rho^2} + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\lambda & \frac{1}{1-\rho^2} + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\lambda & \frac{1}{1-\rho^2} + \left(1 + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\right)\lambda \end{bmatrix}$$

for $s = 0$

$$\frac{1}{\sigma^2}\mathbf{E}\mathbf{v}_t\mathbf{v}_{t+s} = \left(\frac{1}{1-\rho^2} + \left(1 + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\right)\lambda\right) J + \lambda I \tag{55}$$