Notes on estimation of the emissions model (1/27/2021)

This version corrects an error in the formula for the variance. It also proposes an alternative (simpler) method for constructing the variance-covariance matrix $\mathbf{V}$.

1   Summary of model

Emissions in region i at time t equal

$$e_{it} = \frac{1}{b}(b_{0i} + \beta x_{it}) + \frac{1}{b}(\rho\nu_{t-1} + \alpha_t + \mu_{it}) \tag{1}$$

$$\text{with } \nu_{it} = \rho\nu_{t-1} + \alpha_t + \mu_{it} \tag{2}$$

and

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

The aggregate shock is

$$\nu_t \equiv \frac{\sum_i \nu_{it}}{n} = \rho\nu_{t-1} + \alpha_t + \theta_t \text{ with } \theta_t \equiv \frac{\sum_i \mu_{it}}{n} \tag{4}$$

$$\nu_t = \rho\nu_{t-1} + \eta_t \text{ with } \eta_t \equiv \alpha_t + \theta_t \tag{5}$$

2   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 $\kappa(s) = 1$).]

1

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

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

(8)

Dividing by b^2 I have

$$\frac{\mathbf{E}\nu_{i,t}\nu_{j,t+s}}{b^2} = \sigma^2\left[\rho^s\frac{1}{1-\rho^2} + \left((1-\kappa(s))\,\iota(i,j) + \kappa(s)\frac{\rho^s}{n} + \frac{\rho^{2+s}}{n}\frac{1}{1-\rho^2} \right)\lambda \right]$$

(9)

$$\text{with } \sigma^2 \equiv \frac{\sigma_\alpha^2}{b^2} \text{ and } \lambda \equiv \frac{\sigma_\mu^2}{\sigma_\alpha^2}.$$

<mark>[Note that the formula for the covariance has changed.]</mark>

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

For <mark>future use, note that</mark>

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

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

(10)

and

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

(11)

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

<mark>Remember the definitions of ι and κ in equations 6 and 7.</mark>

3   Write the panel as a system of equations

Define

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

(12)

2

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.

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

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

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

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

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.

4   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

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

3

(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).

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

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

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

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

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

$$K =$$

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

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

$$\begin{bmatrix} 3.\,131\,578\,947\lambda + 5.\,263\,157\,895 & 2.\,131\,578\,947\lambda + 5.\,263\,157\,895 & 2.\,368\,421\,053\lambda + 4.\,736\,842\,105 \\ 2.\,131\,578\,947\lambda + 5.\,263\,157\,895 & 3.\,131\,578\,947\lambda + 5.\,263\,157\,895 & 2.\,368\,421\,053\lambda + 4.\,736\,842\,105 \\ 2.\,368\,421\,053\lambda + 4.\,736\,842\,105 & 2.\,368\,421\,053\lambda + 4.\,736\,842\,105 & 3.\,131\,578\,947\lambda + 5.\,263\,157\,895 \\ 2.\,368\,421\,053\lambda + 4.\,736\,842\,105 & 2.\,368\,421\,053\lambda + 4.\,736\,842\,105 & 2.\,131\,578\,947\lambda + 5.\,263\,157\,895 \\ 2.\,131\,578\,947\lambda + 4.\,263\,157\,895 & 2.\,131\,578\,947\lambda + 4.\,263\,157\,895 & 2.\,368\,421\,053\lambda + 4.\,736\,842\,105 \\ 2.\,131\,578\,947\lambda + 4.\,263\,157\,895 & 2.\,131\,578\,947\lambda + 4.\,263\,157\,895 & 2.\,368\,421\,053\lambda + 4.\,736\,842\,105 \end{bmatrix}$$

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

$$\begin{bmatrix}
12.\,778\,947\,37 & 10.\,378\,947\,37 & 10.\,421\,052\,63 & 10.\,421\,052\,63 & 9.\,378\,947\,368 & 9.\,378\,947\,368 \\
10.\,378\,947\,37 & 12.\,778\,947\,37 & 10.\,421\,052\,63 & 10.\,421\,052\,63 & 9.\,378\,947\,368 & 9.\,378\,947\,368 \\
10.\,421\,052\,63 & 10.\,421\,052\,63 & 12.\,778\,947\,37 & 10.\,378\,947\,37 & 10.\,421\,052\,63 & 10.\,421\,052\,63 \\
10.\,421\,052\,63 & 10.\,421\,052\,63 & 10.\,378\,947\,37 & 12.\,778\,947\,37 & 10.\,421\,052\,63 & 10.\,421\,052\,63 \\
9.\,378\,947\,368 & 9.\,378\,947\,368 & 10.\,421\,052\,63 & 10.\,421\,052\,63 & 12.\,778\,947\,37 & 10.\,378\,947\,37 \\
9.\,378\,947\,368 & 9.\,378\,947\,368 & 10.\,421\,052\,63 & 10.\,421\,052\,63 & 10.\,378\,947\,37 & 12.\,778\,947\,37
\end{bmatrix}$$

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

For $s = 0$ I have the block

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

$$\begin{bmatrix}
3.\,131\,578\,947\lambda + 5.\,263\,157\,895 & 2.\,131\,578\,947\lambda + 5.\,263\,157\,895 \\
2.\,131\,578\,947\lambda + 5.\,263\,157\,895 & 3.\,131\,578\,947\lambda + 5.\,263\,157\,895
\end{bmatrix} \tag{20}$$

For $s = 1$ I have the block

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

$$\begin{bmatrix}
2.\,368\,421\,053\lambda + 4.\,736\,842\,105 & 2.\,368\,421\,053\lambda + 4.\,736\,842\,105 \\
2.\,368\,421\,053\lambda + 4.\,736\,842\,105 & 2.\,368\,421\,053\lambda + 4.\,736\,842\,105
\end{bmatrix} \tag{21}$$

For $s = 2$ I have the block

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

$$\begin{bmatrix}
2.\,131\,578\,947\lambda + 4.\,263\,157\,895 & 2.\,131\,578\,947\lambda + 4.\,263\,157\,895 \\
2.\,131\,578\,947\lambda + 4.\,263\,157\,895 & 2.\,131\,578\,947\lambda + 4.\,263\,157\,895
\end{bmatrix} \tag{22}$$

5

5   The old procedure for constructing V

Before I realized that $\mathbf{V}$ 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.

I want a procedure for associating an arbitrary element of the vector $\mathbf{v}$ 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

$$\mathbf{E}v_m v_p = \mathbf{E}v_{m-4(t-1),t}v_{p-4(\tau-1),\tau} \text{ with } t = \left\lfloor \frac{m-1}{4} \right\rfloor + 1 \text{ and } \tau = \left\lfloor \frac{p-1}{4} \right\rfloor + 1 \tag{23}$$

(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 23, I have

$$\tau - t = \left\lfloor \frac{p-1}{4} \right\rfloor + 1 - \left(\left\lfloor \frac{m-1}{4} \right\rfloor + 1 \right) = \left\lfloor \frac{p-1}{4} \right\rfloor - \left\lfloor \frac{m-1}{4} \right\rfloor. \tag{24}$$

Therefore,

$$s \equiv |\tau - t| = \left| \left\lfloor \frac{p-1}{4} \right\rfloor - \left\lfloor \frac{m-1}{4} \right\rfloor \right| \tag{25}$$

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

$(6-1)\,4+2=22$ 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.

6   The 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

$$\mathbf{X}'\Gamma\left(\mathbf{e}-\mathbf{X}\boldsymbol{\beta}\right)\ =\ 0 \tag{26}$$

$$-\frac{4T}{2\sigma^2}+\frac{1}{2\sigma^4}\left(\mathbf{e}-\mathbf{x}\boldsymbol{\beta}\right)'\Gamma\left(\mathbf{e}-\mathbf{x}\boldsymbol{\beta}\right)\ =\ 0. \tag{27}$$

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

$$\frac{1}{2\sigma^2}\left(\sigma^2\mathbf{V}-\tilde{\mathbf{v}}\tilde{\mathbf{v}}'\right)=0. \tag{28}$$

Greene notes that the covariance matrix must be restricted in some way, i.e. it must be possible to write the covariance matrix as $\mathbf{V}=\mathbf{V}\left(\gamma\right)$, where γ is a vector of parameters. For my problem, $\gamma=\left(\rho,\lambda\right)$. (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 26 and 27 to obtain estimates of β and σ^2. Using these, solve equation 28 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 28 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.

$$L = "terms" - \frac{1}{2\sigma^2}\tilde{\mathbf{v}}'\mathbf{\Gamma}\tilde{\mathbf{v}} + \frac{1}{2}\ln|\Gamma| \tag{29}$$

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

$$\frac{dA^{-1}}{d\rho} = -A^{-1}\frac{dA}{d\rho}A^{-1}. \tag{30}$$

Consider the FOC for ρ. I have

$$\frac{\partial \tilde{\mathbf{v}}'\mathbf{\Gamma}\tilde{\mathbf{v}}}{\partial\rho} = \frac{\partial \tilde{\mathbf{v}}'\mathbf{V}^{-1}\tilde{\mathbf{v}}}{\partial\rho} = -\tilde{\mathbf{v}}'\mathbf{V}^{-1}\frac{d\mathbf{V}}{d\rho}\mathbf{V}^{-1}\tilde{\mathbf{v}} \tag{31}$$

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

$$\partial\left(\ln|\Gamma|\right) = Tr\left(\Gamma^{-1}\partial\Gamma\right) \Rightarrow \frac{\partial\left(\ln|\Gamma|\right)}{\partial\rho} = Tr\left(\Gamma^{-1}\frac{\partial\Gamma}{\partial\rho}\right) \tag{32}$$

$$= -Tr\left(\mathbf{V}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\rho}\mathbf{V}^{-1}\right) = -Tr\left(\frac{\partial\mathbf{V}}{\partial\rho}\mathbf{V}^{-1}\right) \tag{33}$$

Using these two equations I have the FOC for ρ:

$$\frac{d\ln L}{d\rho} = \frac{1}{2\sigma^2}\tilde{\mathbf{v}}'\mathbf{V}^{-1}\frac{d\mathbf{V}}{d\rho}\mathbf{V}^{-1}\tilde{\mathbf{v}} - Tr\left(\frac{d\mathbf{V}}{d\rho}\mathbf{V}^{-1}\right) = 0 \tag{34}$$

I also have the FOC for λ

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

Conditional on estimates of σ^2 and β, equations 34 and 34 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 $a =$

(ρ, λ), 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 34 and the second element is given by the left side of equation 35. (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 :

$$\hat{\beta} = \left(\mathbf{X}'V^{-1}\mathbf{X}\right)^{-1} \mathbf{X}'V^{-1}\mathbf{e} \qquad (36)$$

$$\hat{\sigma}^2 = \frac{1}{n} \left(\mathbf{e} - \mathbf{X}'\beta\right)' V^{-1} \left(\mathbf{e} - \mathbf{X}'\beta\right) \qquad (37)$$

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

$$\ln L = -\frac{n}{2} \left[\ln \left(2\pi + \ln \sigma^2\right)\right] - \frac{1}{2\sigma^2} \tilde{\mathbf{v}}'V^{-1}\tilde{\mathbf{v}} + \frac{1}{2} \ln \left|V^{-1}\right| \qquad (38)$$

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 $\| a_j - a_{j-1} \|$ 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.

xxxxxxxxxxxxxxxxxxxxxxxxxxxx
Write it out in detail

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

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

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

$$\frac{1}{\sigma^2}\mathbf{E}\mathbf{v}_t\mathbf{v}_{t+s} = \left(\frac{1}{1-\rho^2} + \left(1 + \frac{\rho^2}{n}\frac{1}{1-\rho^2}\right)\lambda\right)J + \lambda I \tag{39}$$