1 Revised note for Aaron

This note begins with a summary of the old approach, where I factored out σ^2. Next I show how the formulation changes with the "new approach", where I do not factor out σ^2. The following subsection describes my proposed algorithm. I begin this subsection by reviewing the "problem" that Sunny encountered: an estimation procedure that wanted to drive the estimate of σ_α^2 (or in his case, λ, the ratio of variances) to zero. I use the new formulation to show why this problem could *possibly* arise. Then I use a bit of analysis to explain why I am "virtually certain" that the potential problem does *not* in fact arise. Finally, I discuss the implementation of the algorithm.

1.1 Summary of old approach

When Sunny tried to estimate the model using MLE, I factored out $\sigma^2 \equiv \frac{\sigma_\alpha^2}{b^2}$ from the covariance matrix and defined $\lambda \equiv \frac{\sigma_\mu^2}{\sigma_\alpha^2}$. In the paper I showed that

$$\begin{gathered}\frac{\mathbf{E}\nu_{i,t}\nu_{j,t+s}}{b^2} = \frac{\sigma_\alpha^2}{b^2}\left[\rho^s \frac{1}{1-\rho^2} + \left(\left(1-\kappa\left(s\right)\right)\iota\left(i,j\right) + \kappa\left(s\right)\frac{\rho^s}{n} + \frac{\rho^{2+s}}{n}\frac{1}{1-\rho^2}\right)\lambda\right] \\ \text{with } \lambda \equiv \frac{\sigma_\mu^2}{\sigma_\alpha^2},\ \iota\left(i,j\right) = \begin{cases} 1 \text{ for } i = j \\ 0 \text{ for } i \neq j \end{cases},\ \text{and } \kappa\left(s\right) = \begin{cases} 0 \text{ if } s = 0 \\ 1 \text{ if } s \neq 0 \end{cases}.\end{gathered} \tag{1}$$

Using this fact I wrote the the upper triangular part of the covariance matrix $\mathbf{V}$ as

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

(Multiplying both sides by σ^2 gives the expectation of the outer product of errors as $\sigma^2\mathbf{V}$.) Each of the blocks $\mathbf{E}\mathbf{v}_t\mathbf{v}_{t+s}'$ has a simple structure. Denote I_n as the $n$dimensional identity matrix and denote J as the $n \times n$ matrix consisting entirely of 1's. Using equation 1) we obtain

for $s = 0$

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

For $s > 0$

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

1.2 The new approach

The only change I want to make is to NOT factor out σ^2. This change is equivalent to setting $\sigma^2 = 1$ and replacing equation 1 with

$$\begin{aligned}\tfrac{\mathbf{E}\nu_{i,t}\nu_{j,t+s}}{b^2} &= \left[\tfrac{\sigma_\alpha^2}{b^2}\rho^s\tfrac{1}{1-\rho^2} + \left((1-\kappa(s))\,\iota(i,j) + \kappa(s)\tfrac{\rho^s}{n} + \tfrac{\rho^{2+s}}{n}\tfrac{1}{1-\rho^2}\right)\lambda\tfrac{\sigma_\alpha^2}{b^2}\right] \\ &= \tfrac{1}{b^2}\left[\sigma_\alpha^2\rho^s\tfrac{1}{1-\rho^2} + \left((1-\kappa(s))\,\iota(i,j) + \kappa(s)\tfrac{\rho^s}{n} + \tfrac{\rho^{2+s}}{n}\tfrac{1}{1-\rho^2}\right)\sigma_\mu^2\right] \\ &\text{with } \iota(i,j) = \begin{cases} 1 \text{ for } i = j \\ 0 \text{ for } i \neq j \end{cases}, \text{ and } \kappa(s) = \begin{cases} 0 \text{ if } s = 0 \\ 1 \text{ if } s \neq 0 \end{cases}.\end{aligned} \tag{2}$$

The middle line shows the factor $\frac{1}{b^2}$ outside the brackets. However, we are not able to identify b, so for the estimation we just set it equal to 1. I retain it here only so that I don't get confused when I look at these notes in the future.

The following example is adapted from the final section of the note "Statement of Problem"; everything is the same except that now I am not factoring out σ^2. Using the definition

$$\chi(\rho; i, j, s) \equiv \left((1-\kappa(s))\,\iota(i,j) + \kappa(s)\frac{\rho^s}{n} + \frac{\rho^{2+s}}{n}\frac{1}{1-\rho^2}\right), \tag{3}$$

I can write the covariance as

$$\mathbf{E}\nu_{i,t}\nu_{j,t+s} = \frac{1}{b^2}\left[\sigma_\alpha^2\rho^s\frac{1}{1-\rho^2} + \chi(\rho; i, j, s)\,\sigma_\mu^2\right]. \tag{4}$$

(Again, ignore the nuisance parameter $\frac{1}{b^2}$ which is set to 1 in the estimation.)

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 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

$$A = \tfrac{1}{1-\rho^2}\sigma_\alpha^2 + \left(1 + \tfrac{\rho^2}{n(1-\rho^2)}\right)\sigma_\mu^2$$

$$B = \tfrac{1}{1-\rho^2}\sigma_\alpha^2 + \left(\tfrac{\rho^2}{n(1-\rho^2)}\right)\sigma_\mu^2$$

$$C = \tfrac{\rho}{1-\rho^2}\sigma_\alpha^2 + \left(\tfrac{\rho}{n} + \tfrac{\rho^3}{n(1-\rho^2)}\right)\sigma_\mu^2$$

$$E = \tfrac{\rho^2}{1-\rho^2}\sigma_\alpha^2 + \left(\tfrac{\rho^2}{n} + \tfrac{\rho^4}{n(1-\rho^2)}\right)\sigma_\mu^2 = \rho C$$

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

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

For the special case $\rho = 0$, $C = 0$. In this case, $\mathbf{V}$ shows that there is correlation between errors in different regions in the same period, but not between any regions in different periods. For the other special case where $\sigma_\mu^2 = 0$, i.e. where there are no region-specific shocks, the shocks in different regions and the same period are perfectly correlated. For this special case we have

$$A = \tfrac{1}{1-\rho^2}\sigma_\alpha^2$$

$$B = \tfrac{1}{1-\rho^2}\sigma_\alpha^2$$

$$C = \tfrac{\rho}{1-\rho^2}\sigma_\alpha^2$$

$$E = \tfrac{\rho^2}{1-\rho^2}\sigma_\alpha^2 = \rho C$$

$$\mathbf{V} = \frac{1}{1-\rho^2}\sigma_\alpha^2 \begin{bmatrix} 1 & 1 & \rho & \rho & \rho^2 & \rho^2 \\ 1 & 1 & \rho & \rho & \rho^2 & \rho^2 \\ \rho & \rho & 1 & 1 & \rho & \rho \\ \rho & \rho & 1 & 1 & \rho & \rho \\ \rho^2 & \rho^2 & \rho & \rho & 1 & 1 \\ \rho^2 & \rho^2 & \rho & \rho & 1 & 1 \end{bmatrix}.$$

It is easy to show that the determinant of this matrix is 0, i.e. the matrix is singular: the first and second rows are identical, as are the third and fourth, and also the fifth and the sixth. That makes sense: if cross-regional shocks in the same period are perfectly correlated, the covariance matrix is singular. However, for the general case $\sigma_\mu^2 \neq 0$, the matrix is of full rank.[1] The difference is that in this case $A \neq B$.

1.3 My suggested algorithm

1.3.1 The problem that Sunny encountered

When Sunny estimated the model, the estimation procedure drove $\sigma_\mu^2 = 0$, the lower bound. That is, procedure took us to a singular covariance matrix. Sunny might have made a coding error, or there may be something mysterious about factoring out σ_μ^2 that caused the boundary result to arise. In either case, I don't believe that result, for reasons explained below

Using Greene pg 470-471 I wrote the likelihood function as

$$\begin{aligned} \ln L &= -\tfrac{N}{2}\left[\ln\left(2\pi + \ln \sigma^2\right)\right] - \tfrac{1}{2\sigma^2}\mathbf{v}'\mathbf{V}^{-1}\mathbf{v} + \tfrac{1}{2}\ln\left|\mathbf{V}^{-1}\right| \\ &= -\tfrac{N}{2}\left[\ln\left(2\pi + \ln \sigma^2\right)\right] - \tfrac{1}{2\sigma^2}\mathbf{v}'\mathbf{V}^{-1}\mathbf{v} - \tfrac{1}{2}\ln\left|\mathbf{V}\right|. \end{aligned} \tag{5}$$

However, now that we are not factoring σ^2 out of the covariance matrix, the likelihood function is

$$\begin{aligned} \ln L &= -\tfrac{N}{2}\left[\ln\left(2\pi\right)\right] - \tfrac{1}{2}\mathbf{v}'\mathbf{V}^{-1}\mathbf{v} + \tfrac{1}{2}\ln\left|\mathbf{V}^{-1}\right| \\ &= -\tfrac{N}{2}\left[\ln\left(2\pi\right)\right] - \tfrac{1}{2}\mathbf{v}'\mathbf{V}^{-1}\mathbf{v} - \tfrac{1}{2}\ln\left|\mathbf{V}\right|. \end{aligned} \tag{6}$$

I just set $\sigma^2 = 1$ in equation 5. Of course, now we have a different expression for $\mathbf{V}$: it depends on the three parameters $\rho, \sigma_\alpha^2, \sigma_\mu^2$, not (as before) on two parameters ρ, λ. (No free lunches.)

[1] The determinant of $\mathbf{V}$ in this case is $A^6 - 3A^4B^2 - 4A^4C^2\rho^2 - 8A^4C^2 + 8A^3BC^2\rho^2 + 16A^3BC^2 + 16A^3C^3\rho + 3A^2B^4 - 48A^2BC^3\rho - 8AB^3C^2\rho^2 - 16AB^3C^2 + 48AB^2C^3\rho - B^6 + 4B^4C^2\rho^2 + 8B^4C^2 - 16B^3C^3\rho$.

1.3.2 My algorithm

I suggest the following algorithm:

1. Estimate β (the region-specific fixed effects and the common time trend) using OLS, i.e. detrend and demean the data.

2. Using the resulting residuals (i.e. replacing $\mathbf{v}$ with the estimates $\tilde{\mathbf{v}}$ obtained from step 1) choose $\rho, \sigma_\alpha^2, \sigma_\mu^2$ to maximize the likelihood, i.e. to solve
$$\max_{\rho,\sigma_\alpha^2,\sigma_\mu^2} \left[-\frac{1}{2}\tilde{\mathbf{v}}'\mathbf{V}^{-1}\tilde{\mathbf{v}} - \frac{1}{2}\ln|\mathbf{V}| \right]. \tag{7}$$

3. Using the estimates from step 2 form the estimated covariance matrix $\tilde{\mathbf{V}}$ and use this with GLS to obtain new estimate of β and thereby a new estimate of the residuals $\tilde{\mathbf{v}}$. Repeat step 2 and iterate until convergence

1.3.3 Why I think that this problem is well posed

The potential numerical problem is that, independently of $\tilde{\mathbf{v}}$, $-\ln|\mathbf{V}| \to \infty$ as $|\mathbf{V}| \to 0$ and $-\frac{1}{2}\tilde{\mathbf{v}}'\mathbf{V}^{-1}\tilde{\mathbf{v}} \to -\infty$ as $|\mathbf{V}| \to 0$. The two terms go to plus and minus infinity, as $|\mathbf{V}| \to 0$. Sunny found (numerically) that the log term dominated, which meant that the procedure tries to drive $|\mathbf{V}| \to 0$, which in turn means that it wants to drive the estimate of σ_α^2 to zero. (This showed up in Sunny's work as an estimate of $\lambda = 0$.) A bit of manipulation makes me think that this potential numerical problem is not a genuine problem. Here is the manipulation.

Write the adjoint of $\mathbf{V}$ as G, so $\mathbf{V}^{-1} = G \times (|\mathbf{V}|)^{-1}$. Because $\mathbf{V}$ is positive semi-definite, so is G, and $|\mathbf{V}| \geq 0$. For consideration of step 2, we are taking $\tilde{\mathbf{v}}$ as given; it was determined at steps 1 or 3. Recognizing that the adjoint depends on the parameters of the model, we have $G = G\left(\rho, \sigma_\alpha^2, \sigma_\mu^2\right)$. Define $g = g\left(\rho, \sigma_\alpha^2, \sigma_\mu^2\right) \equiv \tilde{\mathbf{v}}'G\left(\rho, \sigma_\alpha^2, \sigma_\mu^2\right)\tilde{\mathbf{v}}$. I am suppressing the dependence of g on $\tilde{\mathbf{v}}$, because here I am taking that vector as given. Also define $\phi\left(\rho, \sigma_\alpha^2, \sigma_\mu^2\right) \equiv \left|\mathbf{V}\left(\rho, \sigma_\alpha^2, \sigma_\mu^2\right)\right|$. Both the determinant and the inner product $\tilde{\mathbf{v}}'G\left(\rho, \sigma_\alpha^2, \sigma_\mu^2\right)\tilde{\mathbf{v}}$ are functions of the parameters of the model.

I have not proven whether or precisely in what manner I can vary ϕ and g independently. The problem is that apart from knowing that $g \geq 0$ and $\phi \geq 0$, I do not know their ranges or precisely how they vary with changes in the parameters. However, for a given numerical value g^* in the range of $g\left(\rho,\sigma_\alpha^2,\sigma_\mu^2\right)$, there is a two-dimensional manifold in $\left(\rho,\sigma_\alpha^2,\sigma_\mu^2\right)$-space satisfying $g^* = g\left(\rho,\sigma_\alpha^2,\sigma_\mu^2\right)$. (I have not actually proven this claim, but it seems pretty "obvious"; it just requires that the derivatives do not vanish – I think!) Moving along that manifold, I obtain different values of $\phi\left(\rho,\sigma_\alpha^2,\sigma_\mu^2\right)$ while $g\left(\rho,\sigma_\alpha^2,\sigma_\mu^2\right) = g^*$ remains constant. Again, I have not proven this claim, but it seems to require only that the derivatives of the two functions are not identical.

In summary, I conjecture that I can indeed vary the two functions independently by varying the parameters of the model. When I say that I can vary the two functions independently, I mean that the value of g, for example, does not uniquely determine the value of ϕ. That is, it is NOT the case that $\phi\left(\rho,\sigma_\alpha^2,\sigma_\mu^2\right) = F\left(g\left(\rho,\sigma_\alpha^2,\sigma_\mu^2\right)\right)$ for some function F. Again, this is only a (strongly held) conjecture. The basis for the conjecture is that both g and ϕ depend – in somewhat different ways – on more than one parameter; in fact they both depend on three parameters. If, to the contrary, they depended on only a single parameter, say x, then when I fix $g(x) = g^*$ I have $x = g^{-1}(g^*)$. In that case, fixing g^* fixes x, thereby fixing $\phi(x) = \phi\left(g^{-1}(g^*)\right)$, i.e. $F \equiv \phi \circ g^{-1}$. In this case, I cannot vary ϕ "independently" as I vary g^*. Thus, the fact that the model has three parameters, not a single parameter, is critical. It is the foundation of my conjecture.

Even if my conjecture (that I can vary g and ϕ independently) is correct, that does not mean that for a given value g^* of $g\left(\rho,\sigma_\alpha^2,\sigma_\mu^2\right)$ I can choose an arbitrary value of ϕ. This limitation is certainly inconvenient, but I don't think that it is fatal, as I now explain.

For the time being, ignore the limitation, and proceed as if we can choose arbitrary ϕ for given g. Using the new notation, I write the maximand as

$$\left[-\frac{1}{2}\tilde{\mathbf{v}}'\mathbf{V}^{-1}\tilde{\mathbf{v}} - \frac{1}{2}\ln|\mathbf{V}|\right] = -\frac{1}{2}\tilde{\mathbf{v}}'G\tilde{\mathbf{v}}\phi^{-1} - \frac{1}{2}\ln\phi = -\frac{1}{2}g\phi^{-1} - \frac{1}{2}\ln\phi.$$

For arbitrary $g > 0$, I have $\frac{d\left(-\frac{1}{2}g\phi^{-1} - \frac{1}{2}\ln\phi\right)}{d\phi} = \frac{1}{2\phi^2}(g-\phi)$. Thus, for $\phi \geq 0$, the maximand is increasing for $\phi < g$ and decreasing for $\phi > g$. It therefore reaches its global max at $\phi = g$. At this point, the maximand equals $\left(-\frac{1}{2}gg^{-1} - \frac{1}{2}\ln g\right) = -\frac{1}{2}\ln g - \frac{1}{2}$. This function is decreasing in g. That is, we maximize the likelihood by making $g \equiv \tilde{\mathbf{v}}'G\left(\rho,\sigma_\alpha^2,\sigma_\mu^2\right)\tilde{\mathbf{v}}$ as small as possible. This is the usual criterion: minimize the residual sum of squares.

Although the above chain of reasoning leads to a sensible conclusion, the chain is based on a dubious premise: the assumption that I can choose *arbitrary* $\phi \geq 0$ for given g. That premise is much stronger than the conjecture that seems to me almost self-evident: the claim that for a given g I can *vary* ϕ. (That is, the choice of g does not "lock in" the choice of ϕ.)

Despite the dubious – and likely false – premise, the analysis is useful. Figure 1 helps explain why. The figure shows the graphs of the likelihood for three values of g, 1, 5 and 10. If the choice of ϕ were truly unconstrained, then we know that the optimal $\phi = g$, as shown above, and as illustrated in the figure. If, as seems more likely, ϕ is constrained in some manner, then the analysis (with the help of the figure) tells us how to select the optimal constrained ϕ. For example, suppose that $g = 1$, and consider three possible types of binding constraints:

- $\phi \leq 0.9$. In this case, we know that the optimal $\phi = 0.9$.
- $\phi \geq 1.1$. In this case, we know that the optimal $\phi = 1.2$.
- $\phi \leq 0.9$ or $\phi \geq 1.1$. In this case, we evaluate the maximand at both boundaries, to find $-\frac{1}{2}1\,(0.9)^{-1} - \frac{1}{2}\ln(0.9) = -0.503$ and $-\frac{1}{2}1\,(1.1)^{-1} - \frac{1}{2}\ln(1.1) = -0.502$, so we know that the optimal constrained value is $\phi = 1.1$.
- Even if we have a more complicated constraint, e.g. one consisting of more than a single gap, we know that we need only consider the boundaries of the feasible region on either side of the unconstrained optimum.

1.3.4 How do we implement the algorithm?

Now that I am reasonably confident that the optimization problem 7 is well-posed (and in particular does not have the "unreasonable" solution $\sigma_\alpha^2 = 0$), we can think more confidently about how to solve it. One possibility is to do a grid search, but it seems more efficient to use some kind of gradient approach. To that end we want the derivatives.

I begin with the derivatives of $\mathbf{V}$ with respect to $\left(\rho, \sigma_\alpha^2, \sigma_\mu^2\right)$. These derivatives use equations 3 and 4 together with the definitions in the last line of

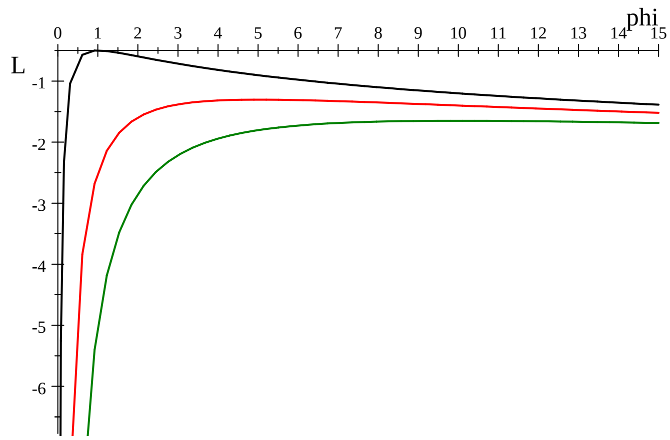

Figure 1: Graphs of the liklihood, L, as a function of ϕ, for $g = 1$ (black), $g = 5$ (red) and $g = 10$ (green). The three graphs reach their maximum at the respective values of g.

equation 2.

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

$$\frac{1}{n}\frac{\rho^{s-1}}{(\rho^2-1)^2}\left(s\kappa+2\rho^2+s\rho^2-s\rho^4-2s\kappa\rho^2+s\kappa\rho^4\right)\sigma_\mu^2+ \tag{8}$$

$$\frac{1}{n}\frac{\rho^{s-1}}{(\rho^2-1)^2}\left(2n\rho^2+ns-ns\rho^2\right)\sigma_\alpha^2$$

and

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

$$\frac{d\left(\sigma_\alpha^2\rho^s\frac{1}{1-\rho^2}+\left((1-\kappa)\iota+\kappa\frac{\rho^s}{n}+\frac{\rho^{2+s}}{n}\frac{1}{1-\rho^2}\right)\sigma_\mu^2\right)}{d\sigma_\mu^2}=\frac{\rho^s}{n}\left(\left(1-\kappa\right)\iota+\kappa+\rho^2\frac{1}{1-\rho^2}\right) \tag{9}$$

This information makes it possible to construct the matrices of derivatives $\frac{d\mathbf{V}}{d\rho}$, $\frac{d\mathbf{V}}{d\sigma_\mu^2}$, and $\frac{d\mathbf{V}}{d\sigma_\alpha^2}$.

Copying and pasting from the "Notes for Aaron" document, I have

$$\frac{\partial\tilde{\mathbf{v}}'\mathbf{V}^{-1}\tilde{\mathbf{v}}}{\partial y}=-\tilde{\mathbf{v}}'\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial y}\mathbf{V}^{-1}\tilde{\mathbf{v}},$$

where y is a placeholder for the parameters $y \in \left\{\rho, \sigma_\alpha^2, \sigma_\mu^2\right\}$. I also have (using Rule 10 in MatrixCookbook.pdf)

$$\frac{\partial \ln |\mathbf{V}|}{\partial y} = Tr\left[V^{-1}\frac{\partial \mathbf{V}}{\partial y}\right].$$

With these intermediate results we can write the gradient of the maximand with respect to $y \in \left\{\rho, \sigma_\alpha^2, \sigma_\mu^2\right\}$ as

$$\frac{\partial\left[-\frac{1}{2}\tilde{\mathbf{v}}'\mathbf{V}^{-1}\tilde{\mathbf{v}} - \frac{1}{2}\ln|\mathbf{V}|\right]}{\partial y} = -\frac{1}{2}\left[-\tilde{\mathbf{v}}'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial y}\mathbf{V}^{-1}\tilde{\mathbf{v}} - Tr\left(V^{-1}\frac{\partial \mathbf{V}}{\partial y}\right)\right].$$

Given $\tilde{\mathbf{v}}$ and values of $\left(\rho, \sigma_\alpha^2, \sigma_\mu^2\right)$ we can numerically calculate values of these derivatives.