Solution to Exercise 10.21

⋆10.21 The GNR proposed in Section 7.8 for NLS estimation of the model (10.86) can be written schematically as

$$\begin{bmatrix} (1-\rho^2)^{1/2}u_1(\boldsymbol{\beta}) \\ u_t(\boldsymbol{\beta}) - \rho u_{t-1}(\boldsymbol{\beta}) \end{bmatrix} = \begin{bmatrix} (1-\rho^2)^{1/2}\boldsymbol{X}_1 & 0 \\ \boldsymbol{X}_t - \rho\boldsymbol{X}_{t-1} & u_{t-1}(\boldsymbol{\beta}) \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ b_\rho \end{bmatrix} + \text{residuals},$$

where $u_t(\boldsymbol{\beta}) \equiv y_t - \boldsymbol{X}_t\boldsymbol{\beta}$ for $t = 1, \ldots, n$, and the last $n-1$ rows of the artificial variables are indicated by their typical elements. Append one extra artificial observation to this artificial regression. For this observation, the regressand is $\big((1-\rho^2)u_1^2(\boldsymbol{\beta})/\sigma_\varepsilon - \sigma_\varepsilon\big)/\sqrt{2}$, the regressor in the column corresponding to ρ is $\rho\sigma_\varepsilon\sqrt{2}/(1-\rho^2)$, and the regressors in the columns correspond to the elements of $\boldsymbol{\beta}$ are all 0. Show that, if at each iteration σ_ε^2 is updated by the formula

$$\sigma_\varepsilon^2 = \frac{1}{n}\Big((1-\rho^2)u_1^2(\boldsymbol{\beta}) + \sum_{t=2}^{n}\big(u_t(\boldsymbol{\beta}) - \rho u_{t-1}(\boldsymbol{\beta})\big)^2\Big),$$

then, if the iterations defined by the augmented artificial regression converge, the resulting parameter estimates satisfy the estimating equations (10.90) that define the ML estimator.

The odd-looking factors of $\sqrt{2}$ in the extra observation are there for a reason: Show that, when the artificial regression has converged, σ_ε^{-2} times the matrix of cross-products of the regressors is equivalent to the block of the information matrix that corresponds to $\boldsymbol{\beta}$ and ρ evaluated at the ML estimates. Explain why this means that we can use the OLS covariance matrix from the artificial regression to estimate the covariance matrix of $\hat{\boldsymbol{\beta}}$ and $\hat{\rho}$.

With the extra observation appended, the regressand of the artificial regression becomes

$$\begin{bmatrix} (1-\rho^2)^{1/2}u_1(\boldsymbol{\beta}) \\ u_t(\boldsymbol{\beta}) - \rho u_{t-1}(\boldsymbol{\beta}) \\ \big((1-\rho^2)u_1^2(\boldsymbol{\beta})/\sigma_\varepsilon - \sigma_\varepsilon\big)/\sqrt{2} \end{bmatrix},$$

and the matrix of regressors becomes

$$= \begin{bmatrix} (1-\rho^2)^{1/2}\boldsymbol{X}_1 & 0 \\ \boldsymbol{X}_t - \rho\boldsymbol{X}_{t-1} & u_{t-1}(\boldsymbol{\beta}) \\ 0 & \rho\sigma_\varepsilon\sqrt{2}/(1-\rho^2) \end{bmatrix}.$$

Remember that the middle "row" in these two matrix expressions actually represents $n-1$ rows, for $t = 2, \ldots, n$. The transpose of the first k columns of the regressor matrix multiplied by the regressand is

$$(1-\rho^2)\boldsymbol{X}_1^\top u_1(\boldsymbol{\beta}) + \sum_{t=2}^{n}(\boldsymbol{X}_t - \rho\boldsymbol{X}_{t-1})^\top\big(u_t(\boldsymbol{\beta}) - \rho u_{t-1}(\boldsymbol{\beta})\big).$$

When this quantity is evaluated at $\hat{\rho}$ and $\hat{\boldsymbol{\beta}}$, it is equal to the left-hand side of the first of equations (10.90). Similarly, the last column of the regressor matrix multiplied by the regressand is

$$\sum_{t=2}^{n} u_{t-1}(\boldsymbol{\beta})\big(u_t(\boldsymbol{\beta}) - \rho u_{t-1}(\boldsymbol{\beta})\big) + \rho u_1^2(\boldsymbol{\beta}) - \frac{\rho\sigma_\varepsilon^2}{1-\rho^2}.$$

When this quantity is evaluated at $\hat{\rho}$ and $\hat{\boldsymbol{\beta}}$, it is equal to the left-hand side of the second of equations (10.90). The formula for updating σ_ε ensures that, when $\rho = \hat{\rho}$ and $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, the third of equations (10.90) is satisfied. Therefore, we conclude that, if the iterations defined by the augmented artificial regression converge, they must converge to parameter estimates $\hat{\rho}$ and $\hat{\boldsymbol{\beta}}$ that satisfy equations (10.90).

The upper left $k \times k$ block of the matrix of cross-products of the regressors, that is, the block that corresponds to $\boldsymbol{\beta}$, is

$$(1-\rho^2)\boldsymbol{X}_1^\top\boldsymbol{X}_1 + \sum_{t=2}^{n}(\boldsymbol{X}_t - \rho\boldsymbol{X}_{t-1})^\top(\boldsymbol{X}_t - \rho\boldsymbol{X}_{t-1}). \tag{S10.51}$$

The lower right 1×1 block, which corresponds to ρ, is

$$\sum_{t=2}^{n} u_{t-1}^2(\boldsymbol{\beta}) + \frac{2\rho^2\sigma_\varepsilon^2}{(1-\rho^2)^2}. \tag{S10.52}$$

The top right off-diagonal block is

$$\sum_{t=2}^{n}(\boldsymbol{X}_t - \rho\boldsymbol{X}_{t-1})^\top u_{t-1}(\boldsymbol{\beta}), \tag{S10.53}$$

and the lower left block is the transpose of this.

We now turn our attention to the information matrix. Minus the matrix of second derivatives of the loglikelihood function (10.89) with respect to $\boldsymbol{\beta}$ is

$$\frac{1}{\sigma_\varepsilon^2}\Big((1-\rho^2)\boldsymbol{X}_1^\top\boldsymbol{X}_1 + \sum_{t=2}^{n}(\boldsymbol{X}_t - \rho\boldsymbol{X}_{t-1})^\top(\boldsymbol{X}_t - \rho\boldsymbol{X}_{t-1})\Big). \tag{S10.54}$$

Since this matrix does not depend on the y_t, there are no expectations to take, and so it is equal to the upper left $k \times k$ block of the information matrix. Clearly the matrix (S10.51) is just (S10.54) multiplied by σ_ε^2. Thus we have the first part of the required result.

Minus the second derivative of the loglikelihood function (10.89) with respect to the scalar parameter ρ is

$$\frac{1}{\sigma_\varepsilon^2}\sum_{t=2}^{n} u_{t-1}^2(\boldsymbol{\beta}) + \frac{1+\rho^2}{(1-\rho^2)^2} - \frac{u_1^2(\boldsymbol{\beta})}{\sigma_\varepsilon^2}. \tag{S10.55}$$

The first term here is the first term of (S10.52) divided by σ_ε^2. The contribution to the information matrix made by the last two terms of (S10.55) is the expectation of the sum of these terms for parameters $\boldsymbol{\beta}$, ρ, and σ_ε^2. Since the expectation of $u_1^2(\boldsymbol{\beta})$ is $\sigma_\varepsilon^2/(1-\rho^2)$, this contribution is

$$\frac{1+\rho^2}{(1-\rho^2)^2} - \frac{1}{1-\rho^2} = \frac{2\rho^2}{(1-\rho^2)^2},$$

which is the second term of expression (S10.52) divided by σ_ε^2.

Finally, the cross-partial derivatives of the loglikelihood function with respect to ρ and $\boldsymbol{\beta}$ are given by the negatives of the elements of the column vector

$$\begin{aligned} &\frac{2\rho \boldsymbol{X}_1^{\top} u_1(\boldsymbol{\beta})}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\varepsilon^2}\sum_{t=2}^{n} \boldsymbol{X}_{t-1}^{\top}\big(u_t(\boldsymbol{\beta}) - \rho u_{t-1}(\boldsymbol{\beta})\big) \\ &\quad + \frac{1}{\sigma_\varepsilon^2}\sum_{t=2}^{n}(\boldsymbol{X}_t - \rho \boldsymbol{X}_{t-1})u_{t-1}(\boldsymbol{\beta}). \end{aligned} \tag{S10.56}$$

For true parameters $\boldsymbol{\beta}$, ρ, and σ_ε^2, the expectation of the first term above is zero. For the second term, note that

$$u_t(\boldsymbol{\beta}) - \rho u_{t-1}(\boldsymbol{\beta}) = u_t - \rho u_{t-1} = \varepsilon_t,$$

and so the expectation of the second term is also zero. The third term is just the vector (S10.53) divided by σ_ε^2.

It may seem odd that we did not take the expectation of the first sum in expression (S10.55) or of the last sum in (S10.56). But notice that the t^{th} term in each of these sums is predetermined at time t, since these terms depend on the dependent variable only through the lagged residual $u_{t-1}(\boldsymbol{\beta})$, or possibly through lags of the dependent variable if there are any of these in the vector $\boldsymbol{X}_t$ of predetermined explanatory variables. The terms of which we did take expectations, on the other hand, depend on current residuals. Thus, except for the factor of $1/\sigma_\varepsilon^2$, the cross-product matrix from the artificial regression is a sum of contributions, each of which can be written generically for observation t as

$$\mathrm{E}_{\boldsymbol{\theta}}\big(\boldsymbol{G}_t^{\top}(\boldsymbol{y},\boldsymbol{\theta})\boldsymbol{G}_t(\boldsymbol{y},\boldsymbol{\theta}) \mid \boldsymbol{y}^{t-1}\big).$$

As can be seen from the definition (10.31), the unconditional expectations of these contributions are just the contributions $\boldsymbol{I}_t(\boldsymbol{\theta})$ that are used to compute the information matrix. It follows that $1/n$ times the cross-product matrix of the artificial regression is a consistent estimator of σ_ε^2 times the asymptotic information matrix $\mathcal{I}(\boldsymbol{\theta})$ defined in equation (10.32).

This result implies that the OLS covariance matrix from the artificial regression can be used to estimate the covariance matrix of $\hat{\boldsymbol{\beta}}$ and $\hat{\rho}$. The OLS covariance matrix is the cross-product matrix multiplied by

$$\frac{n}{n-k}\hat{\sigma}_\varepsilon^2 + \frac{1}{2(n-k)}\left(\frac{(1-\rho^2)u_1^2(\boldsymbol{\beta})}{\sigma_\varepsilon} - \sigma_\varepsilon\right)^2, \tag{S10.57}$$

which is the OLS estimate of the variance from the artificial regression when it has converged. Note that there are $n+1$ artificial observations and $k+1$ artificial parameters. The first term of expression (S10.57) tends to σ_ε^2 as $n \to \infty$, and the second term tends to 0. Thus this expression provides the missing factor of σ_ε^2 needed to estimate the inverse of the information matrix.