where $\hat{\boldsymbol{X}}^*$ denotes the $n \times k$ matrix of the derivatives of the vector of nonlinear functions $\boldsymbol{x}^*(\boldsymbol{\beta}, \rho)$, defined in (10.46), with respect to the elements of $\boldsymbol{\beta}$, evaluated at $(\hat{\boldsymbol{\beta}}, \hat{\rho})$, and

$$\hat{\boldsymbol{V}}(\hat{\rho}, \hat{\omega}) = \begin{bmatrix} \dfrac{n}{1-\hat{\rho}^2} + \dfrac{3\hat{\rho}^2 - 1}{(1-\hat{\rho}^2)^2} & \dfrac{2\hat{\rho}}{\hat{\omega}(1-\hat{\rho}^2)} \\ \dfrac{2\hat{\rho}}{\hat{\omega}(1-\hat{\rho}^2)} & \dfrac{2n}{\hat{\omega}^2} \end{bmatrix}^{-1}.$$

The estimated covariance matrix (10.54) is block-diagonal between $\boldsymbol{\beta}$ and ρ and between $\boldsymbol{\beta}$ and ω (recall that we have ruled out lagged dependent variables). However, unlike the situation with regression models, it is not block-diagonal between ρ and ω. The off-diagonal terms in the (ρ, ω) block of the information matrix are $O(1)$, while the diagonal terms are $O(n)$. Thus $\boldsymbol{V}(\hat{\boldsymbol{\beta}}, \hat{\rho}, \hat{\omega})$ will be asymptotically block-diagonal between $\boldsymbol{\beta}$, ρ, and ω. This is what we would expect, since it is only the first observation, which is asymptotically negligible, that prevents (10.54) from being block-diagonal in the first place.

It is an excellent exercise to derive the estimated covariance matrix (10.54). One starts by taking the second derivatives of (10.51) with respect to all of the parameters of the model to find the Hessian, then takes expectations of minus it to obtain the information matrix. One then replaces parameters by their ML estimates and inverts the information matrix to obtain (10.54). Although this exercise is straightforward, there are plenty of opportunities to make mistakes. For example, Beach and MacKinnon (1978a) fail to take all possible expectations and, as a result, end up with an excessively complicated estimated covariance matrix.

The preceding discussion makes it clear that taking the first observation into account is significantly harder than ignoring it. Even if an appropriate computer program is available, so that estimation is straightforward, one runs into trouble when one wants to test the model. Since the transformed model is no longer a regression model, the Gauss-Newton regression no longer applies and cannot be used to do model specification tests; see Sections 10.8 and 10.9. One could of course estimate the model twice, once taking account of the first observation, in order to obtain the most efficient possible estimates, and once dropping it, in order to be able to test the specification, but this clearly involves some extra work. The obvious question that arises, then, is whether the additional trouble of taking the first observation into account is worth it.

There is a large literature on this subject, including Kadiyala (1968), Rao and Griliches (1969), Maeshiro (1976, 1979), Beach and MacKinnon (1978a), Chipman (1979), Spitzer (1979), Park and Mitchell (1980), Ansley and Newbold (1980), Poirier (1978a), Magee (1987), and Thornton (1987). In many cases, retaining the first observation yields more efficient estimates but not by very much. However, when the sample size is modest and there is one or