93.4.5. Nonlinear Testing and Forecasting Asymptotics with Potential Rank Failure—Solution, proposed by Peter C.B. Phillips.

(a) The maximum likelihood estimator \( \hat{\theta} \) is obtained by applying nonlinear least squares to equation (1), i.e., by minimizing the sum of squares so that

\[
\sum_{t=1}^{T} (y_t - \alpha x_{1t} - \beta x_{2t} - \alpha \beta x_{3t})^2 = \sum_{i=1}^{T} (y_i - x_i' \gamma(\theta))^2, \text{ say.}
\]

Under the given assumptions, we have the limit theory

\[\sqrt{T}(\hat{\theta} - \theta) \rightarrow_d N(0, V_\theta),\]

where

\[V_\theta = \sigma^2 \left[ \frac{\partial \gamma(\theta)'}{\partial \theta} M \frac{\partial \gamma(\theta)}{\partial \theta'} \right]^{-1}\] and \[M = \lim_{T \rightarrow \infty} (T^{-1} X' X) - I_3.\]
Because
\[ \frac{\partial \gamma(\theta')}{\partial \theta} = \begin{bmatrix} 1 & 0 & \beta \\ 0 & 1 & \alpha \end{bmatrix}, \]
we obtain
\[ V_\theta = \sigma^2 \begin{bmatrix} 1 + \beta^2 & \alpha \beta \\ \alpha \beta & 1 + \alpha^2 \end{bmatrix}^{-1} = \frac{\sigma^2}{1 + \alpha^2 + \beta^2} \begin{bmatrix} 1 + \alpha^2 & -\alpha \beta \\ -\alpha \beta & 1 + \beta^2 \end{bmatrix}. \]

(b) Let \( \Psi(\theta) = \alpha \beta \). Then the Wald statistic for testing \( H_0: \Psi(\theta) = 0 \) is
\[ W_T = T \Psi(\dot{\theta}) \left[ \frac{\partial \gamma(\dot{\theta})}{\partial \theta} \right] \left( V_\theta \frac{\partial \Psi(\dot{\theta})}{\partial \theta} \right)^{-1} \Psi(\dot{\theta}), \]
where
\[ \Psi(\dot{\theta}) = \hat{\alpha} \hat{\beta}, \]
\[ V_\theta = \hat{\sigma}^2 \left[ \frac{\partial \gamma(\dot{\theta})}{\partial \theta} M_{xx} \frac{\partial \gamma(\dot{\theta})}{\partial \theta} \right]^{-1}, \]
\[ \frac{\partial \Psi(\dot{\theta})}{\partial \theta} = (\hat{\beta}, \hat{\alpha}), \]
and \( \hat{\sigma}^2 = T^{-1} \Sigma_T (y_T - x_T \gamma(\dot{\theta}))^2 \) is the MLE of the error variance \( \sigma^2 \) in the model.

Because \( M_{xx} \rightarrow I_3 \) as \( T \rightarrow \infty \), the limit distribution of \( W_T \) is the same as that of the statistic where \( M_{xx} \) is replaced by \( I_3 \), namely
\[ \hat{W}_T = \frac{T(\hat{\alpha} \hat{\beta})^2(1 + \hat{\alpha}^2 + \hat{\beta}^2)}{\hat{\sigma}^2(\hat{\alpha}^2 + \hat{\beta}^2)}. \quad (\ast) \]

Now, under \( H_0 \) we have \( \alpha \beta = 0 \) and
\[ \sqrt{T} \hat{\alpha} \hat{\beta} \rightarrow_d N \left( 0, \frac{\partial \Psi(\theta)}{\partial \theta} \right) V_\theta \frac{\partial \Psi(\theta)}{\partial \theta} \right) \equiv N(0, \sigma^2(\alpha^2 + \beta^2)/(1 + \alpha^2 + \beta^2)). \quad (\ast\ast) \]

Because \( (\hat{\alpha}, \hat{\beta}) \rightarrow_p (\alpha, \beta) \), it follows from (\ast) and (\ast\ast) that
\[ W_T \hat{W}_T \rightarrow_d \chi_1^2, \]
except when \( \alpha = \beta = 0 \). Note that in this case (where \( \alpha \beta = 0 \) and \( H_0 \) is satisfied), \( \dot{\theta}' = (\hat{\alpha}, \hat{\beta}) \rightarrow_p (0, 0) \) and
\[ \sqrt{T} \dot{\theta} \rightarrow_d N(0, \sigma^2 I_2). \quad (\dagger) \]
Thus, \( W_T, \tilde{W}_T \) have the same limit distribution as

\[
\frac{T (\hat{\alpha} \hat{\beta})^2}{\sigma^2 (\hat{\alpha}^2 + \hat{\beta}^2)} \sim \frac{\{(T^{1/2} \hat{\alpha} / \hat{\sigma}) (T^{1/2} \hat{\beta} / \hat{\sigma})\}^2}{\{(\hat{\alpha} / \hat{\sigma})^2 + (\hat{\beta} / \hat{\sigma})^2\}} \rightarrow_d \chi^2_\alpha \chi^2_\beta / (\chi^2_\alpha + \chi^2_\beta),
\]

where \( \chi^2_\alpha \) and \( \chi^2_\beta \) are both \( \chi^2_1 \) (chi-square with one degree of freedom) variates and are statistically independent in view of (††).

(c) The forecast \( \hat{y}_{T+1} \) is

\[
\hat{y}_{T+1} = \hat{\alpha} x_{1T+1} + \hat{\beta} x_{2T+1} + \hat{\alpha} \hat{\beta} x_{3T+1},
\]

and forecast error is

\[
y_{T+1} - \hat{y}_{T+1} = u_{T+1} + (\alpha - \hat{\alpha}) x_{1T+1} + (\beta - \hat{\beta}) x_{2T+1} + (\alpha \hat{\beta} - \hat{\alpha} \hat{\beta}) x_{3T+1},
\]

which is asymptotically equivalent to

\[
u_{T+1} + x_{T+1}^T (\partial \gamma(\theta) / \partial \theta') (\hat{\theta} - \theta) = u_{T+1} + \tilde{x}_{T+1}^T (\hat{\theta} - \theta), \text{ say}.
\]

The asymptotic variance of the forecast error is then

\[
\text{var}(y_{T+1} - \hat{y}_{T+1}) = \hat{\sigma}^2 \{1 + \hat{x}_{T+1}^T (X'X)^{-1} \hat{x}_{T+1}\},
\]

where

\[
\hat{x}_{T+1} = x_{T+1} \frac{\partial \gamma(\hat{\theta}) / \partial \theta'}{\partial \theta} = (x_{1T+1} + \hat{\beta} x_{2T+1}, x_{2T+1} + \hat{\alpha} x_{3T+1}),
\]

and

\[
\hat{X}' \hat{X} = \Sigma_{\hat{X}' \hat{X}} = \frac{\partial \gamma(\hat{\theta}) / \partial \theta'}{\partial \theta} (X'X) \frac{\partial \gamma(\hat{\theta})}{\partial \theta'}.
\]

(d) When \( \theta = 0 \) (i.e., \( \alpha = \beta = 0 \)), the model is simply \( y_t = u_t \) i.i.d. \( N(0, \sigma^2 I) \). Note that the null hypothesis

\[
H_0: \alpha \beta = 0
\]

holds in this case. We now have

\[
\sqrt{T} (\hat{\theta} - \theta) = \sqrt{T} \hat{\theta} \rightarrow_d N(0, \sigma^2 I_2),
\]

\[
W_T \rightarrow_d \chi^2_\alpha \chi^2_\beta / (\chi^2_\alpha + \chi^2_\beta)
\]

as in (††), and

\[
\text{var}(y_{T+1} - \hat{y}_{T+1}) = \hat{\sigma}^2 \{1 + \hat{x}_{T+1}^T (X'X)^{-1} \hat{x}_{T+1}\},
\]

as in (†††).