Econometric Theory, **20**, 2004, 989–993. Printed in the United States of America.
DOI: 10.1017/S0266466604205096

PROBLEMS AND SOLUTIONS

SOLUTIONS

03.5.1. A Concise Derivation of the Wallace and Hussain Fixed Effects Transformation[1]—Solution

Badi H. Baltagi (the poser of the problem)
Texas A&M University

In vector form the disturbances can be written as

$$u = Z_\mu \mu + Z_\lambda \lambda + \nu,$$

where $Z_\mu = I_N \otimes \iota_T$, I_N is an identity of dimension T, and ι_N is a vector of ones dimension N, $Z_\lambda = \iota_N \otimes I_T$, μ is of dimension $N \times 1$, λ is of dimension $T \times 1$, and ν is of dimension $NT \times 1$. In general, if $\Delta = [X_1, X_2]$ then $P_\Delta = P_{X_1} + P_{[Q_{X_1} X_2]}$, where $P_\Delta = \Delta(\Delta'\Delta)^{-1}\Delta'$ denotes the projection matrix on Δ and $Q_\Delta = I - P_\Delta$. Applying this result to $\Delta = [Z_\mu, Z_\lambda]$, one gets

$$P_\Delta = P_{Z_\mu} + P_{[Q_{[Z_\mu]} Z_\lambda]} = P + P_{QZ_\lambda} = P + QZ_\lambda(Z'_\lambda Q Z_\lambda)^{-1} Z'_\lambda Q,$$

where $P = I_N \otimes \bar{J}_T$ with $\bar{J}_T = \iota_T \iota'_T / T$ and $Q = I_N \otimes E_T$ with $E_T = I_T - \bar{J}_T$. Using the fact that $QZ_\lambda = \iota_N \otimes E_T$, $Z'_\lambda Q Z_\lambda = NE_T$, $(Z'_\lambda Q Z_\lambda)^- = (1/N)E_T$, one gets $P_{QZ_\lambda} = \bar{J}_N \otimes E_T$. Hence

$$P_\Delta = P + \bar{J}_N \otimes E_T,$$

which means that

$$\begin{aligned} Q_\Delta &= I_{NT} - P_\Delta = Q - \bar{J}_N \otimes E_T \\ &= I_N \otimes E_T - \bar{J}_N \otimes E_T \\ &= E_N \otimes E_T \end{aligned}$$

as required. Here Q_Δ is the *fixed effects transformation* derived by Wallace and Hussain (1969). Note that the order does not matter; i.e., one could have orthogonalized on Z_λ.

NOTE

1. An excellent solution has been independently proposed by Francisco J. Goerlich, University of Valencia.

REFERENCE

Wallace, T.D. & A. Hussain (1969) The use of error components models in combining cross-section and time-series data. *Econometrica* 37, 55–72.

 0266-4666/04 $12.00

03.5.2. Consistent Standard Errors for Target Variance Approach to GARCH Estimation—Solution

Dennis Kristensen and Oliver Linton (the posers of the problem)
London School of Economics

The problem falls in the framework of two-step generalized method of moments estimators (GMM estimators) as described in Newey and McFadden (1994, Sec. 6). Their general results may be applied; here, however, we give a direct derivation of the asymptotic variance. For simplicity, assume that consistency of $\hat{\theta}$ has already been proved. We choose our parameter space as $\Theta = \{\theta | \beta + \gamma < 1\}$ to ensure that the second moment exists. Let θ_0 and σ_0^2 denote the true parameter values and let $\mathcal{N}(\theta_0)$ and $\mathcal{N}(\sigma_0^2)$, respectively, denote (shrinking) neighborhoods of these.

By a standard Taylor expansion,

$$0 = \frac{\partial \ell_T(\hat{\theta}, \hat{\sigma}^2)}{\partial \theta} = \frac{\partial \ell_T(\theta_0, \hat{\sigma}^2)}{\partial \theta} + \frac{\partial^2 \ell_T(\bar{\theta}, \hat{\sigma}^2)}{\partial \theta \partial \theta'} (\hat{\theta} - \theta_0)$$

for some $\bar{\theta} \in [\hat{\theta}, \theta_0]$ and

$$\frac{\partial \ell_T(\theta_0, \hat{\sigma}^2)}{\partial \theta} = \frac{\partial \ell_T(\theta_0, \sigma_0^2)}{\partial \theta} + \frac{\partial^2 \ell_T(\theta_0, \bar{\sigma}^2)}{\partial \theta \partial \sigma^2} (\hat{\sigma}^2 - \sigma_0^2)$$

for some $\bar{\sigma}^2 \in [\hat{\sigma}^2, \sigma_0^2]$. If

$$\text{(i)} \quad \begin{bmatrix} \sqrt{T} \dfrac{\partial \ell_T(\theta_0, \sigma_0^2)}{\partial \theta} \\ \sqrt{T}(\hat{\sigma}^2 - \sigma_0^2) \end{bmatrix} \Rightarrow N\left(0, \begin{bmatrix} \Sigma_{\theta\theta} & \Sigma_{\theta\sigma} \\ \Sigma_{\sigma\theta} & \Sigma_{\sigma\sigma} \end{bmatrix}\right)$$

and

$$\text{(ii)} \quad \frac{\partial^2 \ell_T(\bar{\theta}, \hat{\sigma}^2)}{\partial \theta \partial \theta'} \xrightarrow{P} H_{\theta\theta}, \qquad \text{(iii)} \quad \frac{\partial^2 \ell_T(\theta_0, \bar{\sigma}^2)}{\partial \theta \partial \sigma^2} \xrightarrow{P} H_{\theta\sigma},$$

we have that

$$\sqrt{T}(\hat{\theta} - \theta_0) \Rightarrow N(0, H_{\theta\theta}^{-1}(\Sigma_{\theta\theta} + H_{\theta\sigma}\Sigma_{\sigma\theta} + \Sigma_{\theta\sigma}H_{\sigma\theta} + H_{\theta\sigma}\Sigma_{\sigma\sigma}H_{\sigma\theta})H_{\theta\theta}^{-1}). \quad \textbf{(2)}$$

If ε_t are normally distributed, var$[\hat{\theta}]$ is larger than the variance of the full maximum likelihood estimator (MLE) of θ_0, but in the absence of normality the comparison could go either way. If σ^2 were known instead of estimated, then the asymptotic variance would simplify to $H_{\theta\theta}^{-1}\Sigma_{\theta\theta}H_{\theta\theta}^{-1}$.

In the following we show that (i)–(iii) hold: First, derive the first and second derivatives of ℓ_T:

$$\frac{\partial \ell_T(\theta,\sigma^2)}{\partial \theta} = \frac{1}{2T}\sum_{t=1}^{T}\frac{\partial \log \sigma_t^2}{\partial \theta}\left(\frac{y_t^2}{\sigma_t^2}-1\right), \tag{3}$$

$$\frac{\partial^2 \ell_T(\theta,\sigma^2)}{\partial \theta \partial \theta'} = \frac{1}{2T}\sum_{t=1}^{T}\frac{\partial^2 \log \sigma_t^2}{\partial \theta \partial \theta'}\left(\frac{y_t^2}{\sigma_t^2}-1\right) - \frac{1}{2T}\sum_{t=1}^{T}\frac{\partial \log \sigma_t^2}{\partial \theta}\frac{\partial \log \sigma_t^2}{\partial \theta'}\frac{y_t^2}{\sigma_t^2}, \tag{4}$$

$$\frac{\partial^2 \ell_T(\theta,\sigma^2)}{\partial \theta \partial \sigma^2} = \frac{1}{2T}\sum_{t=1}^{T}\frac{\partial^2 \log \sigma_t^2}{\partial \theta \partial \sigma^2}\left(\frac{y_t^2}{\sigma_t^2}-1\right) - \frac{1}{2T}\sum_{t=1}^{T}\frac{\partial \log \sigma_t^2}{\partial \theta}\frac{\partial \log \sigma_t^2}{\partial \sigma^2}\frac{y_t^2}{\sigma_t^2}, \tag{5}$$

where $\sigma_t^2 = \sigma_t^2(\theta,\sigma^2)$. The derivative of $\log \sigma_t^2$ with respect to $\alpha = (\theta,\sigma^2)$ is given by $\partial \log \sigma_t^2/\partial\alpha = \partial\sigma_t^2/\partial\alpha \cdot \sigma_t^{-2}$ where

$$\frac{\partial \sigma_t^2}{\partial \beta} = -\sigma^2 + \sigma_t^2 + \beta\frac{\partial \sigma_{t-1}^2}{\partial \beta}, \qquad \frac{\partial \sigma_t^2}{\partial \gamma} = -\sigma^2 + y_{t-1}^2 + \beta\frac{\partial \sigma_{t-1}^2}{\partial \gamma},$$

$$\frac{\partial \sigma_t^2}{\partial \sigma^2} = 1 + \beta\frac{\partial \sigma_{t-1}^2}{\partial \sigma^2}.$$

Iterating the preceding expressions yields

$$\frac{\partial \sigma_t^2}{\partial \beta} = -\sigma^2\frac{1-\beta^{t-1}}{1-\beta} + \sum_{s=0}^{t-1}\beta^s\sigma_{t-1-s}^2,$$

$$\frac{\partial \sigma_t^2}{\partial \gamma} = -\sigma^2\frac{1-\beta^{t-1}}{1-\beta} + \sum_{s=0}^{t-1}\beta^s y_{t-1-s}^2, \qquad \frac{\partial \sigma_t^2}{\partial \sigma^2} = \frac{1-\beta^{t-1}}{1-\beta},$$

where we have taken $\sigma_t^2|_{t=0}$ to be given. From these expressions, one can check that (see Lee and Hansen, 1994, Lemmas 8 and 10)

$$\left|\frac{\partial \log \sigma_t^2}{\partial \theta}\right|, \left|\frac{\partial \log \sigma_t^2}{\partial \sigma^2}\right| \le C_1 \tag{6}$$

uniformly over $(\theta,\sigma^2) \in \mathcal{N}(\theta_0) \times \mathcal{N}(\sigma_0^2)$ for some constant $C_1 < \infty$. Also, there exists $C_2 < \infty$ such that

$$E\left[\frac{y_t^2}{\sigma_t^2}\right] \le \frac{1}{\sigma^2(1-\gamma-\beta)}E[y_t^2] \le C_2 E[y_t^2] < \infty \tag{7}$$

uniformly over $(\theta,\sigma^2) \in \mathcal{N}(\theta_0) \times \mathcal{N}(\sigma_0^2)$. This proves that (4) and (5) both are uniformly bounded by functions with finite expectations. They are furthermore continuous in (θ,σ^2), so by standard results concerning uniform convergence (see, e.g., Tauchen, 1985) (ii) and (iii) follow.

To show (i), we first observe that under (1), we have that $\{y_t, \sigma_t^2(\theta_0,\sigma_0^2)\}$ is β-mixing with exponentially decaying mixing coefficients (cf. Carrasco and Chen, 2002). A standard central limit theorem (CLT) for mixing sequences may therefore be applied to obtain the desired result given that Σ exists. Note that the weak convergence of $\sqrt{T}(\partial \ell_T(\theta_0,\sigma_0^2)/\partial\theta)$ alone can be proved using martingale arguments, but $\hat\sigma^2 - \sigma_0^2$ is not a martingale so we have to appeal to CLT for mixing sequences instead. The asymptotic variance matrix is given by

$$\Sigma_{\theta\theta} = E\left[\left(\frac{\partial \ell(\theta_0,\sigma_0^2)}{\partial\theta}\right)\left(\frac{\partial \ell(\theta_0,\sigma_0^2)}{\partial\theta}\right)'\right]$$

and

$$\Sigma_{\sigma\sigma} = \operatorname{var}(y_t^2) + 2\sum_{s=1}^{\infty} \operatorname{cov}(y_0^2, y_s^2), \qquad \Sigma_{\sigma\theta} = E\left[\left(\frac{\partial \ell_T(\theta_0,\sigma_0^2)}{\partial\theta}\right)' \hat\sigma^2\right].$$

Using the inequalities established earlier, it is easily seen that $\Sigma_{\theta\theta}$ is well defined if $E[\varepsilon_t^4] < \infty$. But for $\Sigma_{\sigma\sigma}$ to be finite we must require $E[y_t^4] < \infty$. A necessary and sufficient condition for this is $\nu_4 \equiv E[\varepsilon_t^4] < \infty$ and

$$\nu_4\gamma^2 + 2\gamma\beta + \beta^2 < 1$$

(cf. He and Teräsvirta, 1999). This is a stronger condition than (1). In effect, we need to restrict our parameter space Θ further to obtain asymptotic normality.

As a result of the correlation structure, explicit expressions for Σ will require tedious and rather lengthy algebra, and the resulting expressions will most likely be very complicated. But we are still able to derive a simple estimator of the asymptotic variance: we have already found consistent estimators of $H_{\theta\theta}$ and $H_{\theta\sigma}$, so we only need to find an estimator of Σ. Here, we use the general covariance estimator proposed by Newey and West (1987) and check that their conditions are satisfied in our case. Define the function

$$m_t(\theta,\sigma^2) = \left[\frac{\partial \log \sigma_t^2}{\partial\theta}\left(\frac{y_t^2}{\sigma_t^2} - 1\right), (y_t^2 - \sigma^2)\right]'$$

that satisfies

$$\operatorname{Var}\left(\frac{1}{T}\sum_{i=1}^{T} m_t(\theta,\sigma^2)\right) = \Sigma = \begin{pmatrix} \Sigma_{\theta\theta} & \Sigma_{\theta\sigma} \\ \Sigma_{\sigma\theta} & \Sigma_{\sigma\sigma} \end{pmatrix}.$$

We then apply the conditions of Newey and West (1987, Theorem 2) on m, which are as follows: (i) There exists a function $\bar m$ such that $\|m_t(\theta,\sigma^2)\| \leq$

https://doi.org/10.1017/S0266466604215092 Published online by Cambridge University Press

$\bar{m}(y_t, y_{t-1})$ uniformly over $(\theta, \sigma^2) \in \mathcal{N}(\theta_0) \times \mathcal{N}(\sigma_0^2)$ and $E[\bar{m}(y_t, y_{t-1})^2] < \infty$; (ii) $E[\|m_t(\theta_0, \sigma_0^2)\|^{4(1+\delta)}] < \infty$ for some $\delta > 0$; and (iii) $\{y_t\}$ is ϕ-mixing with mixing coefficients of size $2r/(2r-1)$ for some $r > 1$. If these are satisfied, we may choose

$$\hat{\Sigma} = \hat{\Omega}_0 + \sum_{i=1}^{N_T} w_j(N_T)\hat{\Omega}_j,$$

$$\hat{\Omega}_j = \frac{1}{T}\sum_{t=1}^{T} m_t(\hat{\theta}, \hat{\sigma}^2) m_{t-j}(\hat{\theta}, \hat{\sigma}^2)'$$

as an estimator of the variance of Σ where $w_j(N_T)$ are weights and N_T is an increasing sequence. Under certain conditions on $w_j(N_T)$ and N_T (see Newey and West, 1987, p. 705), $\hat{\Sigma}$ is consistent. By the inequalities established in (6) and (7) together with the assumption that $E[y_t^4] < \infty$, $\|m_t(\theta, \sigma^2)\| \le \bar{m}_t$ uniformly over $(\theta, \sigma^2) \in \mathcal{N}(\theta_0) \times \mathcal{N}(\sigma_0^2)$ for some random variable with $E[\bar{m}_t^2] < \infty$, which proves (i). If $E[y_t^{8(1+\delta)}] < \infty$ for some $\delta > 0$ then (ii) is satisfied. Finally, (iii) holds by the aforementioned result of Carrasco and Chen (2002).

We conclude that if $E[y_t^4] < \infty$, we have asymptotic normality of $\hat{\theta}$; if furthermore $E[y_t^{8(1+\delta)}] < \infty$, we may estimate its asymptotic variance by

$$\mathrm{AsVar}\{\sqrt{T}(\hat{\theta} - \theta_0)\} \doteq \hat{H}_{\theta\theta}^{-1}\hat{V}\hat{H}_{\theta\theta}^{-1},$$

where

$$\hat{V} = \hat{\Sigma}_{\theta\theta} + \hat{H}_{\theta\sigma}\hat{\Sigma}_{\sigma\theta} + \hat{\Sigma}_{\theta\sigma}\hat{H}_{\sigma\theta} + \hat{H}_{\theta\sigma}\hat{\Sigma}_{\sigma\sigma}\hat{H}_{\sigma\theta},$$

$$\hat{H}_{\theta\theta} = \frac{\partial^2 \ell_T(\hat{\theta}, \hat{\sigma}^2)}{\partial\theta\partial\theta'}, \qquad \hat{H}_{\theta\sigma} = \frac{\partial^2 \ell_T(\hat{\theta}, \hat{\sigma}^2)}{\partial\theta\partial\sigma^2},$$

and with $\hat{\Sigma}$ given previously.

REFERENCES

Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.

He, C. & T. Teräsvirta (1999) Fourth moment structure of the GARCH(p,q) process. *Econometric Theory* 15, 824–846.

Lee, S.-W. & B. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–53.

Newey, W.K. & D.L. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, chap. 36. Elsevier.

Newey, W.K. & K.D. West (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Tauchen, G.E. (1985) Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* 30, 415–443.