Adaptive Estimation of Cointegrating Regressions with ARMA Errors

Hodgson, Douglas J.

Working Paper No. 408

September 1995

Adaptive Estimation of Cointegrating Regressions with ARMA Errors

Douglas J. Hodgson

Rochester Center for Economic Research
Working Paper No. 408

September 1995

ADAPTIVE ESTIMATION OF COINTEGRATING REGRESSIONS WITH ARMA ERRORS

Douglas J. Hodgson[1]

Department of Economics

University of Rochester

Rochester, NY

14627-0156

June 12, 1995

Adaptive maximum likelihood estimators are derived for the parameters of a cointegrating regression whose errors follow a stationary and invertible ARMA process with innovations of unknown distribution. It is shown how to use preliminary consistent estimates of these innovations to nonparametrically estimate their density, which can in turn be used to construct an asymptotically efficient iterative estimator of the cointegrating vector. The asymptotic distribution of this estimator is derived and its efficiency gains relative to the Gaussian pseudo-MLE are evaluated for the case of Student t innovations.

JEL no. C22; Keywords - nonnormality, kernel estimation, kurtosis, asymptotic efficiency

---

[1] This paper is based on Chapter 1 of my Ph.D. thesis at Yale University. I am grateful to Peter Phillips and Oliver Linton for advice and assistance, and to the Alfred P. Sloan Foundation and the Social Sciences and Humanities Research Council of Canada for financial assistance.

I. INTRODUCTION

In many econometric models, maximum likelihood estimators (MLE's) are used because of their appealing properties of asymptotic normality and asymptotic efficiency. However, use of ML techniques requires the specification of the underlying density function generating the data. In many cases, it is appropriate to use the analytically tractable normal distribution as an approximation to the true distribution and compute the Gaussian MLE. However, in many important situations, the Gaussian assumption is highly unrealistic and can lead to the use of very inefficient estimates. In economics, an important example of such a situation is the analysis of financial time series such as stock prices or exchange rates, whose generating processes are well documented to have excessively thick tails (see, for example, Mandelbrot (1963), Fama (1963, 1965), Mittnik and Rachev (1993) and McGuirk, Robertson, and Spanos (1993)).

In such estimation problems, where a Gaussian MLE is inappropriate, adaptive estimation provides a highly attractive alternative. Adaptive estimation is employed when the underlying density function of the data generating process is of unknown shape. An adaptive estimator shares the asymptotic optimality properties of an MLE, differing from the latter in that a nonparametric estimate of the score function of the log-likelihood replaces the analytic expression for the score that would be used if the density were known. An adaptive estimator can be viewed as an MLE when the shape of the likelihood is unknown. A simulation study by McDonald and White (1993) finds that adaptive estimators compare quite favourably with OLS, LAD, GMM, and M-estimators in the estimation of a (non-cointegrating) non-Gaussian linear regression model.

Adaptive estimators can be expected to lead to considerable efficiency gains in the estimation of many cointegrated models, especially in the areas of financial economics and international finance. This is because the estimation methods most commonly employed in applied cointegration studies rely on Gaussianity assumptions in order to claim optimality properties. Optimal estimation of cointegrated models in the Gaussian case is extensively analyzed by Phillips (1991). The estimation of single-equation models using least squares techniques can be expected to give inefficient and misleading estimates when applied to thick-tailed time series such as speculative prices. That this is indeed the case has been illustrated by Phillips (1993) in the context of the estimation of exchange rate models.

The present paper is concerned with the adaptive estimation of single equation cointegrating regressions whose errors follow a stationary and invertible ARMA process. It draws on previous work by Kreiss (1987b), Steigerwald (1989,1992), and, especially, Jeganathan (1994). Kreiss (1987b) derives adaptive estimators of stationary and invertible ARMA processes, and many of his results are directly applicable to the adaptive estimation of the ARMA component of the model considered here. Steigerwald (1989,1992) analyzes linear regressions with an ARMA error process, but the regression is not a cointegrating one. Jeganathan (1994) provides a very comprehensive review of adaptive estimation in time series models and introduces adaptive estimation of cointegrated models. The latter analysis deals with cointegrating regressions whose errors are iid. The present analysis extends Jeganathan's by drawing on results of Kreiss (1987b) and Steigerwald (1989) to allow for ARMA errors.

In section II below, we present the model to be analyzed and introduce some notation. In section III, we compute the limit distribution of the sample log-likelihood ratios, showing that this distribution has a locally asymptotically normal (LAN) component, associated with the parameters of the ARMA process, and a locally asymptotically mixed normal (LAMN) component, associated with the parameters of the cointegrating vector, and that these two components are mutually independent. This independence is important because it implies that if we treat the parameters of the ARMA process as nuisance parameters, with the cointegrating vector being the parameter of interest, the latter is adaptively estimable. In other words, knowledge of the true values of the ARMA parameters will not lead to any asymptotic gains in efficiency in the estimation of the cointegrating vector. In section IV, we obtain an expression for an optimal estimator of the model given the preceding LAN/LAMN results and assuming knowledge of the underlying density of the innovations. In section V, a method of computing an adaptive estimator using nonparametric estimates of the density of the innovation process is given. The asymptotic distribution of the adaptive estimator is derived, and its efficiency gains relative to the Gaussian pseudo-MLE are evaluated. In section VI, we extend the analysis to allow for an intercept in the cointegrating regression.

The following notation is used throughout the paper. I_s denotes the identity matrix of dimension s, $|x|$ the Euclidean norm of the vector x, $I(\cdot)$ the indicator function, $N(x,V)$ the distribution of a random variable that is normal with mean vector x and covariance matrix V, and $MN(x,V)$ a mixed normal distribution, i.e. one in which the covariance matrix V is random. The inequalities $X > Y$ and $X \geq Y$, when applied to matrices, signify that the difference X-Y is positive definite and positive semi-definite, respectively. We simplify the notation by writing $\int_0^1 M$ in place of $\int_0^1 M(r)dr$ when $M(r)$ is a Brownian motion process defined on the interval $[0,1]$. $L(X|P)$ denotes the distribution (or law) of X with respect to the probability measure P. When P is the distribution of X itself, $L(X|P)$ is abbreviated to $L(X)$. The weak convergence of probability measures is denoted by the symbol $\Rightarrow$.

II. THE MODEL AND NOTATION

The observable data consist of the univariate series $\{Y_t\}_{t=1}^n$ and the m-vector series $\{X_t\}_{t=1}^n$. It is assumed that all $m+1$ series are I(1) and that a single cointegrating relationship exists among them. It is further assumed that deviations of the variables from their cointegrating relationship follow a stationary and invertible ARMA(p,q) process. Formally, we have:

(1) $\quad Y_t = BX_t + u_t$

(2) $$u_t = \sum_{j=1}^{p} a_j u_{t-j} + \sum_{j=1}^{q} b_j \varepsilon_{t-j} + \varepsilon_t$$

(3) $$X_t = X_{t-1} + v_t.$$

Furthermore, we assume that:

(i) p and q are known;

(ii) the ARMA process is stationary and invertible, i.e. the polynomials $1 - \sum_{j=1}^{p} a_j z^j$ and $1 - \sum_{j=1}^{q} b_j z^j$ have zeros strictly outside the unit circle, and that these polynomials have no zeros in common;

(iii) the innovations $(\varepsilon_t, v_t')'$ are iid from the unknown elliptically symmetric Lebesgue density $p(\varepsilon, v)$, which has the property that $0 < \lambda^2 = \iint |\psi(\varepsilon,v)|^2 p(\varepsilon,v)dvd\varepsilon < \infty$, where $\psi(\varepsilon,v) = (\partial p(\varepsilon,v)/\partial\varepsilon)/p(\varepsilon,v)$; and

(iv) the initial conditions are $(\varepsilon_{1-q},\ldots,\varepsilon_0; Y_{1-p},\ldots,Y_0; X_{1-p},\ldots,X_0)$. They are assumed to be drawn from the distribution $f_0(\varepsilon_{1-q},\ldots,\varepsilon_0; Y_{1-p},\ldots,Y_0; X_{1-p},\ldots,X_0;\theta_n)$, which has the property that

$$f_0(\varepsilon_{1-q},\ldots,\varepsilon_0; Y_{1-p},\ldots,Y_0; X_{1-p},\ldots,X_0;\theta_n) - f_0(\varepsilon_{1-q},\ldots,\varepsilon_0; Y_{1-p},\ldots,Y_0; X_{1-p},\ldots,X_0;\theta)$$
$$= o_p(1) \ in\, P_{\theta,n}$$

as $\theta_n \to \theta$, where θ_n and θ are parameters defined below.

Remark: In assumption (iii), $\psi(\varepsilon, v)$ denotes the (negative of the) first element of the score vector $p(\varepsilon, v)$, and λ^2 denotes the first element on the diagonal of the information matrix of $p(\varepsilon, v)$. The fact that these quantities are unknown to the investigator is the central problem to be addressed in formulating an adaptive estimator. Our elliptical symmetry assumption implies that $\psi(\varepsilon, v)$ is anti-symmetric in ε, i.e. that $\psi(-\varepsilon, v) = -\psi(\varepsilon, v)$. This property is important in our derivation of an adaptive estimator.

We denote the vector of ARMA coefficients by $\eta = (a_1, \ldots, a_p; b_1, \ldots, b_q)$ and the $p+q+m$-dimensional full parameter vector by $\theta = (\eta, B)$. We let $\{h_n\} = \{(h_{\eta n}', h_{Bn}')'\}$ be a bounded sequence (where $h_n \in R^{p+q+m}, h_{\eta n} \in R^{p+q}, h_{Bn} \in R^m$) and define the scaling matrix $\delta_n = diag\left[n^{-1/2} I_{p+q}, n^{-1} I_m\right]$. We can then write the local representation of the full parameter vector θ as $\theta_n = \theta + \delta_n h_n = (\eta_n, B_n)$. We assume that θ and $\{\theta_n\}$ belong to the parameter space Θ, defined by allowing B to take any value in R^m and η to take any value in R^{p+q} subject to the restrictions mentioned in assumption (ii) above (this also implies restrictions on the possible values of the vectors $h_{\eta n}$). Note that θ_n converges to θ, but does so at different rates in different directions of the parameter space. In directions associated with transitory dynamics, the rate of convergence is $n^{1/2}$, whereas in those associated with nonstationary dynamics, the rate is n.

We use the parameters of the MA component of the model to define the infinite sequence of constants, $\{\gamma_k(\theta)\}$, as follows:

$$(4) \qquad (1+b_1 z+\ldots+b_q z^q)^{-1} = \sum_{k=0}^{\infty} \gamma_k(\theta) z^k,$$

with the following formula holding:

$$\gamma_s(\theta) + b_1\gamma_{s-1}(\theta)+\ldots+b_q\gamma_{s-q}(\theta) = 0 \quad \forall s \geq 1,$$

with $\gamma_s(\theta) = 0 \quad \forall s < 0$ and $\gamma_0(\theta) = 1$. Note that $\gamma_k(\theta) \to 0$ as $k \to \infty$. Using this, we can show that:

$$\begin{aligned}
\varepsilon_t &= u_t - \sum_{i=1}^{p} a_i u_{t-i} + \sum_{k=1}^{t-1} \gamma_k(\theta)(u_{t-k} - \sum_{i=1}^{p} a_i u_{t-k-i}) + \sum_{s=0}^{q-1} \varepsilon_{-s}(\sum_{k=0}^{s} \gamma_{t+s-k}(\theta) a_k) \\
(5) \qquad &= Y_t - BX_t - \sum_{i=1}^{p} a_i (Y_{t-i} - BX_{t-i}) + \sum_{k=1}^{t-1} \gamma_k(\theta)[(Y_{t-k} - BX_{t-k}) - \sum_{i=1}^{p} a_i (Y_{t-k-i} \\
&\quad - BX_{t-k-i})] + \sum_{s=0}^{q-1} \varepsilon_{-s}(\sum_{k=0}^{s} \gamma_{t+s-k}(\theta) a_k).
\end{aligned}$$

This is a standard result giving a formula for the innovations of the ARMA process in terms of observed variables, parameter values, and initial conditions. We now introduce the following $(p+q)$-vector:

$$Z_{t-1}(\eta_n, \eta, B) = \sum_{k=0}^{t-1} \gamma_k(B, \eta_n)(u_{t-1-k}, \ldots, u_{t-p-k}; \varepsilon_{t-1-k}, \ldots, \varepsilon_{t-q-k})'.$$

This vector is important to our theory. It consists only of stationary variables and is used in our derivation of an expression for that part of the score vector of the sample that is associated with the ARMA component of the model. We obtain $\gamma_j(B, \eta_n)$ from (4) by replacing $b_1, \ldots, b_q$ with $b_1^n, \ldots, b_q^n$, where b_j^n is the relevant element of η_n.

In the following, we will sometimes write $\varepsilon_t(\theta) = \varepsilon_t$ to denote the true innovations in the ARMA process (i.e. the innovations as evaluated at the true parameter vector θ). When we write $\varepsilon_t(\theta_1)$, where $\theta_1 \neq \theta$, we refer to the innovations as evaluated at θ_1 when θ is still the true parameter vector. We can then cite the following result, derived by Steigerwald (1989, p. 20) (see also Kreiss (1987b, p. 115) and Jeganathan (1994)):

$$\varepsilon_t(\theta_n) - \varepsilon_t(\theta) = (\theta - \theta_n)' \sum_{j=0}^{t-1} \gamma_j(B, \eta_n)[u_{t-1-j}, \ldots, u_{t-p-j}, \varepsilon_{t-1-j}, \ldots, \varepsilon_{t-q-j}, \tag{6}$$
$$(X_{t-j} - \sum_{k=1}^{p} a_k^n X_{t-j-k})']' .$$

In (6), a_k^n denotes the relevant element of the vector η_n. We can rewrite the right-hand side of (6) as follows:

$$(\eta - \eta_n)' Z_{t-1}(\eta_n, \eta, B) + (B - B_n)' (\sum_{j=0}^{t-1} \gamma_j(B, \eta_n)[X_{t-j} - \sum_{k=1}^{p} a_k^n X_{t-j-k}]).$$

Defining $\Gamma_{t-1}^*(\eta_n, B) = \sum_{j=0}^{t-1} \gamma_j(B, \eta_n)[X_{t-j} - \sum_{k=1}^{p} a_k^n X_{t-j-k}]$ gives us

(7) $\quad \varepsilon_t(\theta_n) - \varepsilon_t(\theta) = (\eta - \eta_n)' Z_{t-1}(\eta_n, \eta, B) + (B - B_n)' \Gamma^*_{t-1}(\eta_n, B).$

We furthermore define $\Gamma_{t-1}(\eta_n, B) = \Gamma^*_{t-1}(\eta_n, B) - v_t$. Note that $\Gamma_{t-1}(\eta_n, B)$ and $\Gamma^*_{t-1}(\eta_n, B)$ are both m-vectors integrated of order one. The subtraction of v_t from $\Gamma^*_{t-1}(\eta_n, B)$ ensures that $\Gamma_{t-1}(\eta_n, B)$ is independent of (ε_t, v_t) and therefore also of $\psi(\varepsilon_t, v_t)$. This property allows our derivation of the LAMN theory for the nonstationary component of the model in the argument following Definition 3.2.

We now place our model within the framework described in Section 4 of Jeganathan (1994), where we let $\underline{Y}_t = (\underline{Y}_0, Y_1, \ldots, Y_t)$, $\underline{Y}_0 = (X_{1-p}, \ldots, X_n; Y_{1-p}, \ldots, Y_0; \varepsilon_{1-q}, \ldots, \varepsilon_0)$, and $A_t = \sigma(\underline{Y}_t)$ be the σ −field generated by $\underline{Y}_t$. We also let $f_0(\underline{Y}_0, \theta)$ be the density of $\underline{Y}_0$ with respect to a σ −finite measure. Our model can then be written[1]:

(8) $\quad Y_t = g_t(\underline{Y}_{t-1}, \theta) + \varepsilon_t,$

where, writing $g_t(\underline{Y}_{t-1}, \theta) = g_{t-1}(\theta)$, we have

$$
\begin{aligned}
g_{t-1}(\theta) = BX_t + \sum_{i=1}^{p} a_i(Y_{t-i} - BX_{t-i}) - \sum_{k=1}^{t-1} \gamma_k(\theta)[(Y_{t-k} - BX_{t-k}) \\
- \sum_{i=1}^{p} a_i(Y_{t-k-i} - BX_{t-k-i})] - \sum_{s=0}^{q-1} \varepsilon_{-s}\left(\sum_{k=o}^{s} \gamma_{t-s-k}(\theta) a_k\right).
\end{aligned}
$$

---

[1] Note that since X_t, and therefore v_t, is present in $g_{t-1}(\theta)$, it is generally the case that $g_{t-1}(\theta)$ is not independent of ε_t, so that our model does not have the non-linear time series structure given by equation (27) of Jeganathan (1994). In our model, the conditional density of Y_t given A_{t-1} is the same as the conditional density of ε_t given v_t.

Since $Y_t = g_{t-1}(\theta) + \varepsilon_t(\theta) = g_{t-1}(\theta_n) + \varepsilon_t(\theta_n)$, we have

$g_{t-1}(\theta_n) - g_{t-1}(\theta) = \varepsilon_t(\theta) - \varepsilon_t(\theta_n) = d_t(\theta_n, \theta)$. From (7), we have

$$
\begin{aligned}
d_t(\theta_n, \theta) &= (\eta_n - \eta)' Z_{t-1} + (B_n - B)' \Gamma^*_{t-1} \\
&= (\theta_n - \theta)' H^*_{t-1} \qquad (9)\\
&= h_n' \delta_n H^*_{t-1}
\end{aligned}
$$

where $H^*_{t-1} = (Z_{t-1}', \Gamma^*_{t-1}')'$. In (9), we have simplified notation by writing $Z_{t-1}(\eta_n, \eta, B) = Z_{t-1}$, with Γ^*_{t-1} and H^*_{t-1} defined analogously. We also define $\Gamma_{t-1} = \Gamma_{t-1}(\eta_n, \eta, B)$, $H_{t-1} = (Z_{t-1}', \Gamma_{t-1}')'$, and $H_{t-1}(\theta) = H_{t-1}(\eta, \eta, B)$. Note that $H_{t-1} = H^*_{t-1} - \begin{bmatrix} 0 \\ v_t \end{bmatrix}$, so that (9) becomes

$$
d_t(\theta_n, \theta) = h_n' \delta_n H_{t-1} + h_n' \delta_n \begin{bmatrix} 0 \\ v_t \end{bmatrix}. \qquad (10)
$$

Equation (10) is fundamental to the theory developed in this paper. The first component on the right-hand side is a linear combination of the elements of the $(p+q+m)-$vector $\delta_n H_{t-1}$. The first $p+q$ elements of this vector are stationary variables scaled by $n^{-1/2}$, while the last m elements are nonstationary variables scaled by n. As will be shown below, multiplying this vector by $\psi(\varepsilon_t, v_t)$ and summing over t will give us an expression for the score vector of the sample, while summing the outer products of $\delta_n H_{t-1} \psi(\varepsilon_t, v_t)$ over t will give us an expression for the information matrix of the sample. The structure of H_{t-1} implies that the first $p+q$ elements of the score are associated with stationary dynamics, with the remaining m elements being associated with nonstationary dynamics. The second component on the right-hand side of (10) turns out to be asymptotically negligible.

III. LAN AND LAMN LIMIT THEORY

In this section, we derive the asymptotic distribution of the log-likelihood ratios of the sample. Let $P_{\theta,n}$ be the distribution of the sample of size n with parameter θ. We seek the asymptotic distribution of the log-likelihood ratio

$$\Lambda_n(\theta_n, \theta) = \log\left(dP_{\theta_n,n} \middle/ dP_{\theta,n} \right).$$

We shall derive, in Theorem 3.1, the probability limit of the log-likelihood ratio, which will be used to show that the component of the model associated with the ARMA parameters is LAN and the component associated with the cointegrating vector is LAMN, and that, furthermore, the two components are asymptotically independent. This independence allows us to adaptively estimate the vector B when η is treated as an unknown nuisance parameter. Formal definitions of LAN and LAMN (and of the important associated property of contiguity) are given below. It is important to show that our model falls within the LAN/LAMN family because characterizations of optimal estimators for the parameters of such models have been derived in the statistics literature (see Fabian and Hannan (1982), LeCam and Yang (1990, pp. 80-88), and Jeganathan (1994)).

The key result in showing that our model falls within the LAN/LAMN family is Theorem 3.1, given below. A more general family of models than the LAN/LAMN is the *locally asymptotically quadratic* (LAQ) family. A model falls into the LAQ family when its sample likelihood ratio, $\Lambda_n(\theta_n, \theta)$, can be asymptotically approximated by a quadratic function of the vector h_n (recall that $h_n = \delta_n^{-1}(\theta_n - \theta)$). Jeganathan (1994) and LeCam and Yang (1990) formally define LAQ families. Theorem 3.1 shows that the likelihood ratios in the present model can be asymptotically approximated by the quadratic given in equation (11).

<u>*Theorem 3.1*</u>: *The likelihood ratios* $\Lambda_n(\theta_n, \theta)$ *have the following quadratic approximation in our model*:

$$
\begin{aligned}
\Lambda_n(\theta_n,\theta) &= -\sum_{t=1}^{n} h_n'\,\delta_n H_{t-1}\psi(\varepsilon_t, v_t) - \frac{\lambda^2}{2} h_n'\,\delta_n \sum_{t=1}^{n} H_{t-1}H_{t-1}'\,\delta_n h_n + o_p(1) \ \ in\ P_{\theta,n} \\
&= -\sum_{t=1}^{n} h_n'\,\delta_n H_{t-1}(\theta)\psi(\varepsilon_t, v_t) - \frac{\lambda^2}{2} h_n'\,\delta_n \sum_{t=1}^{n} H_{t-1}(\theta)H_{t-1}(\theta)'\,\delta_n h_n + o_p(1) \ \ in\ P_{\theta,n}.
\end{aligned} \tag{11}
$$

<u>Remark</u>: Theorem 3.1 shows that the likelihood ratios $\Lambda_n(\theta_n, \theta)$ can be asymptotically approximated by a quadratic function of the vector h_n, where the linear term is the scaled sample score vector and the quadratic term is the scaled sample information matrix, which, as shown by (87) in the Appendix, converges weakly to the asymptotic information matrix $\lambda^2 J(\theta)$, where

$$
J(\theta) = \begin{bmatrix} J_{ZZ}(\theta) & 0 \\ 0 & J_{\Gamma\Gamma}(\theta) \end{bmatrix}.
$$

The non-random $(p+q)$-dimensional submatrix $J_{ZZ}(\theta)$ is defined in (77) and is associated with the ARMA component of the model. The random m-dimensional submatrix $J_{\Gamma\Gamma}(\theta)$ is defined in (78) and is associated with the cointegrating vector. The block diagonality of the information matrix is important because it is a necessary condition for our claims that the cointegrating vector can be efficiently estimated adapting for the ARMA parameters.

We now show how Theorem 3.1 implies the stated LAN/LAMN limit theory for our model. We begin with the following definition (Jeganathan (1994)):

<u>*Definition 3.2*</u>: *The family $\{P_{\theta,n};\theta \in \Theta\}$ is said to have LAMN likelihood ratios at $\theta \in \Theta$ if the quadratic approximation (11) holds and if*

$$
\begin{aligned}
(12)\qquad & L\left(\sum_{t=1}^{n}\delta_n H_{t-1}(\theta)\psi(\varepsilon_t,v_t),\lambda^2\sum_{t=1}^{n}\delta_n H_{t-1}(\theta)H_{t-1}(\theta)'\,\delta_n \middle| P_{\theta,n}\right)\\
& \Rightarrow L\left(S(\theta)^{1/2}N(0,I),S(\theta)\right),
\end{aligned}
$$

where $S(\theta)$ is positive definite almost surely and N(0,I) is a standard Gaussian independent of $S(\theta)$. When $S(\theta)$ is non-random, LAMN likelihoods are called LAN.

For our model, we shall verify that (12) holds and that $S(\theta)=\lambda^2 J(\theta)$, which is block diagonal. The first block, $\lambda^2 J_{ZZ}(\theta)$, is of dimension $p+q$, corresponds to the coefficients of the ARMA process, and is non-random. The second block, $\lambda^2 J_{\Gamma\Gamma}(\theta)$, is of dimension m, corresponds to the cointegrating vector, and is random. Hence, we will show that the first component is LAN, that the second one is LAMN, and that the two are independent.

We begin the verification of (12) by noting that

$$\sum_{t=1}^{n} \delta_n H_{t-1}(\theta)\psi(\varepsilon_t, v_t) = \sum_{t=1}^{n} \begin{bmatrix} \frac{1}{\sqrt{n}} Z_{t-1}(\eta, \eta, B)\psi(\varepsilon_t, v_t) \\ \frac{1}{n}\Gamma_{t-1}(\eta, B)\psi(\varepsilon_t, v_t) \end{bmatrix}. \tag{13}$$

Kreiss (1987b) analyzes the first component of (13), showing that

$$L\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_{t-1}(\eta, \eta, B)\psi(\varepsilon_t, v_t) \Big| P_{n,\theta}\right) \Rightarrow N\left(0, \lambda^2 J_{ZZ}(\theta)\right).$$

As for the second component, we can show that

$$\frac{1}{n}\sum_{t=1}^{n} \Gamma_{t-1}(\eta, B)\psi(\varepsilon_t, v_t) \Rightarrow c_p \gamma(1)\int_0^1 M_2 dM_1 ,$$

where M_1 is a Brownian motion independent of M_2 and with variance λ^2. Since, by Phillips and Park (1988), we have

$$L\left(\int_0^1 M_2 dM_1\right) = MN\left(0, \lambda^2 \int_0^1 M_2 M_2'\right),$$

it follows that

$$L\left(\frac{1}{n}\sum_{t=1}^{n} \Gamma_{t-1}(\eta, B)\psi(\varepsilon_t, v_t) \Big| P_{\theta,n}\right) \Rightarrow MN\left(0, \lambda^2 J_{\Gamma\Gamma}(\theta)\right).$$

To show that (12) holds with $S(\theta) = \lambda^2 J(\theta)$, we must show that

$$(14) \qquad \frac{1}{n^{3/2}} \sum_{t=1}^{n} \Gamma_{t-1}(\eta, B) Z_{t-1}(\eta, \eta, B)' \, \psi^2(\varepsilon_t, \nu_t) = o_p(1) \;\; in \; P_{\theta,n} .$$

The left-hand side of (14) can be rewritten as

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} \Gamma_{t-1}(\eta, B) Z_{t-1}(\eta, \eta, B)' \left(\psi^2(\varepsilon_t, \nu_t) - \lambda^2\right) + \frac{\lambda^2}{n^{3/2}} \sum_{t=1}^{n} \Gamma_{t-1}(\eta, B) Z_{t-1}(\eta, \eta, B)'$$
$$= o_p(1) + o_p(1) = o_p(1) \;\; in \; P_{\theta,n}$$

by arguments similar to those used to show (63).

One important consequence of the LAN/LAMN result derived above is that the sequences of probability measures $\{P_{\theta,n}\}$ and $\{P_{\theta_n,n}\}$ are contiguous, and therefore have the property that the sequence of statistics $\{T_n\}$ is $o_p(1)$ in $P_{\theta_n,n}$ if and only if it is $o_p(1)$ in $P_{\theta,n}$ (see LeCam (1960, p. 40) and LeCam and Yang (1990, p. 20)). We shall use this fact below because we shall compute statistics using residuals $\varepsilon(\theta_n)$ from a consistently estimated model in lieu of the true innovations $\varepsilon(\theta)$, and shall use the fact that the latter statistics are $o_p(1)$ in $P_{\theta,n}$ to show that the former are $o_p(1)$ in $P_{\theta_n,n}$.

IV. <u>CONSTRUCTION OF AC ESTIMATORS</u>

In this section, we are concerned with the construction of efficient estimators for the model described above. Jeganathan (1994) describes a class of estimators termed asymptotically centering (AC) that possess certain optimality properties in

LAN/LAMN models. Since the calculation of such estimators assumes knowledge of the distribution of the density function of the innovation process, which is generally not known, it is not of immediate practical use. However, as will be described in the next section, nonparametric estimates of this density can be used to construct adaptive estimators that share the asymptotic optimality properties of AC estimators.

The following definition of AC estimators is given by Jeganathan (1994):

Definition 4.1: *If the model is LAMN or LAN at θ, we call a sequence $\{\hat{\theta}_n\}$ of estimators AC if*

(15) $$\delta_n^{-1}(\hat{\theta}_n - \theta) - S_n^{-1}(\theta)W_n(\theta) = o_p(1) \text{ in } P_{\theta,n},$$

where

(16) $$S_n(\theta) = \lambda^2 \sum_{t=1}^{n} \delta_n H_{t-1}(\theta) H_{t-1}(\theta)' \delta_n,$$

(17) $$W_n(\theta) = -\sum_{t=1}^{n} \delta_n H_{t-1}(\theta) \psi(\varepsilon_t, v_t).$$

Remark: In Definition 4.1, $W_n(\theta)$ denotes the scaled sample score vector and $S_n(\theta)$ denotes the scaled sample information matrix.

Before proceeding to the construction of AC estimators, we present some notation and assumptions. We assume that consistent estimates $\hat{S}_n$ of $S_n(\theta)$ exist.

Under the assumptions of our model, conditions (C2) and (C5) of Jeganathan (1994) hold ((C2) because Θ is open and $\delta_n \to 0$, and (C5) because δ_n does not depend on θ). Let $\{\theta_n^*\}$ be a sequence of preliminary estimates such that:

$$\delta_n^{-1}(\theta_n^* - \theta) = O_p(1) \;\; in\; P_{\theta,n} \quad \forall\theta \in \Theta.$$

Define θ_n^{**} as a discretized version of θ_n^*. The following definition of discretization is quoted from Jeganathan (1994):

> Partition the space R^{p+q+m} into cubes $C_i, i \geq 1,$ of sides of length unity, and let $C_{ni} = \delta_n C_i = \{\delta_n u: u \in C_i\}$. If $\theta_n^* \in \Theta \cap C_{ni}$, take $\theta_n^{**} = t_{ni}$, where t_{ni} is some fixed point in $\Theta \cap C_{ni}$, which will necessarily be non-empty since $\theta_n^* \in \Theta$. The θ_n^{**} constructed in this way preserves the properties of θ_n^* in the sense that $\theta_n^{**} \in \Theta$ and $\delta_n^{-1}(\theta_n^{**} - \theta) = O_p(1)$ in $P_{\theta,n}$ for all $\theta \in \Theta$.

In practice, any preliminary estimator θ_n^* we may want to use will effectively already be discretized, since we will only compute it to a prespecified finite number of decimal places. Define the quantity $W_n^*(\theta)$ as in Proposition 3 of Jeganathan (1994). (We do not repeat the definition here because it would involve the introduction of a considerable amount of new notation. For our purposes, the important characteristic of

$W_n^*(\theta)$ is the fact (proved in Proposition 3 of Jeganathan(1994)) that

$W_n^*(\theta) = W_n(\theta) - \hat{S}_n h_n + o_p(1)$ *in* $P_{\theta,n}$ for every bounded $\{h_n\}$.)

Given the above definitions and assumptions, we have from Theorem 2 of Jeganathan (1994) that $\hat{\theta}_n$ as given in equation (18) below is an AC estimator:

$$(18) \qquad \hat{\theta}_n = \theta_n^{**} + \delta_n \hat{S}_n^{-1} W_n^*(\theta_n^{**}).$$

The following discussion, based on Jeganathan (1994), shows heuristically why (18) is an AC estimator and why we require the notion of a discretized estimator. As noted above, Proposition 3 of Jeganathan (1994) proves that

$$(19) \qquad W_n^*(\theta_n) = W_n(\theta) - \hat{S}_n h_n + o_p(1) \;\; in \; P_{\theta,n}$$

for every bounded $\{h_n\}$ and for every $\theta \in \Theta$. Defining the estimator

$$(20) \qquad \bar{\theta}_n = \theta_n^* + \delta_n \hat{S}_n^{-1} W_n^*(\theta_n^*),$$

and replacing h_n in (19) with $\delta_n^{-1}(\theta_n^* - \theta)$ (which, recall, is $O_p(1)$ in $P_{\theta,n}$), it would seem that we could combine (19) and (20) to conclude that

$$\delta_n^{-1}(\bar{\theta}_n - \theta) = \hat{S}_n^{-1} W_n(\theta) + o_p(1) \;\; in \; P_{\theta,n},$$

so that $\bar{\theta}_n$ is an AC estimator. However, it is not strictly correct to replace h_n with $\delta_n^{-1}(\theta_n^* - \theta)$ in (19), since $\delta_n^{-1}(\theta_n^* - \theta)$, although confined with probability arbitrarily close to one to a bounded interval, may assume any of an uncountably infinite number of values within such an interval. This would require the replacement of (19) with the stronger condition that

$$\sup_{|h|\le\alpha}\left|W_n^*(\theta_n)-(W_n(\theta)-\hat{S}_n h)\right| = o_p(1) \;\; in\ P_{\theta,n}\ \forall\alpha>0.$$

We can avoid this problem by replacing θ_n^* with θ_n^{**}. For then the quantity $\delta_n^{-1}(\theta_n^{**}-\theta)$ can only assume one of a finite number of points in any bounded interval.

AC estimators are asymptotically equivalent to maximum likelihood estimators, being asymptotically normal (or mixed normal) with covariance matrix equal to the inverse of the Fisher information matrix. They are therefore optimal according to the locally asymptotically minimax criterion (for a discussion, see, for example, Ghosh (1985, pp. 318-320)) and are consequently also known as locally asymptotically minimax estimators.

V. ADAPTIVE ESTIMATION

As mentioned earlier, the construction of the AC estimator given in equation (18) requires us to know the density function of the innovations (ε_t, v_t). In particular, the quantity $\psi(\varepsilon_t, v_t)$ is required. In this section, we show how to calculate an adaptive estimator. Our main problem is to estimate ψ. This is done nonparametrically, using density estimators of the density function $p(\varepsilon, v)$. The density estimator described below is similar to ones used by Bickel (1982), Kreiss (1987b), Jeganathan (1988), and Linton (1993).

Following Jeganathan (1994), we assume that $p(\varepsilon,v)$ is elliptically symmetric, so that

$$p(\varepsilon,v) = |\det\Omega|^{-\frac{1}{2}} f^{*}\left(\left|\Omega^{-\frac{1}{2}}\binom{\varepsilon}{v}\right|\right)$$

for some f^{*}, where $\Omega = \begin{bmatrix}\omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22}\end{bmatrix}$.

Denote the characteristic function of (ε,v) as

$$cf(s) = \phi(s'\Omega s).$$

We then have

$$\operatorname{cov}\binom{\varepsilon_1}{v_1} = k\Omega,$$

where $k = -2\phi'(0)$ (see Fang, Kotz, and Ng (1990, p.43) and Mitchell (1989, pp.290-291)). If $p(\varepsilon,v)$ is Gaussian, then $k=1$ and $\operatorname{cov}(\varepsilon_1, v_1')' = \Omega$.

Define $\ell_{11} = (\omega_{11} - \omega_{21}'\Omega_{22}^{-1}\omega_{21})^{\frac{1}{2}}$, so that

$$p(\varepsilon, v) = |\det \Omega|^{-\frac{1}{2}} f^*\left(\left|\begin{matrix} \ell_{11}^{-1}(\varepsilon - \omega_{21}' \Omega_{22}^{-1} v) \\ \Omega_{22}^{-\frac{1}{2}} v \end{matrix}\right|\right)$$

$$\equiv f(z, v)$$

where $z = \varepsilon - \omega_{21}' \Omega_{22}^{-1} v$.

We have assumed that the density $p(\varepsilon, v)$ is elliptically symmetric. This means that the conditional density of ε given v is symmetric with mean $\omega_{21}' \Omega_{22}^{-1} v$ (see Fang, Kotz, and Ng (1990, p. 45)), and that, given v, f is symmetric about zero in z, so that adaptive estimation following the methods of Bickel (1982) and Kreiss (1987b) can be carried out using nonparametric estimates of the density $f(z,v)$.

In what follows, the construction of a nonparametric estimator of $\psi(\varepsilon_t, v_t)$ is described. We define

$$z_t(\theta) = \varepsilon_t(\theta) - \omega_{21}' \Omega_{22}^{-1} v_t,$$

$$\pi(z, v, \sigma) = \left(1 \Big/ \left(\sigma\sqrt{2\pi}\right)^{m+1} \right) \exp\left(-\left(|z|^2 + |v|^2\right) \Big/ 2\sigma^2 \right),$$

$$\hat{f}_{\sigma,t}(x, y, \theta) = \frac{1}{2(n-1)} \sum_{\substack{i=1 \\ i \neq t}}^{n} \left\{ \pi(x + z_i(\theta), y + v_i, \sigma) + \pi(x - z_i(\theta), y + v_i, \sigma) \right\},$$

and let $\hat{f}'_{\sigma,t}(x,y,\theta)$ be the partial derivative of $\hat{f}_{\sigma,t}$ with respect to x. In this formulation, $\pi(z, v, \sigma)$ is a normal kernel density estimator with smoothing parameter σ; the larger is σ, the smoother is the density estimate. Silverman (1986) gives a detailed description of kernel estimators and the selection of the value of smoothing parameters. Note that, by construction, $\hat{f}$ is symmetric about zero in x and $\hat{f}'$ is anti-symmetric about zero in x. It follows that $\hat{\psi}_{n,t}$, as described below, is also anti-symmetric about zero in x. This fact is important to our development of an adaptive estimator below.

Given fixed y, define

$$\hat{\psi}_{n,t}(x,y,\theta) = \begin{cases} \dfrac{\hat{f}'_{\sigma(n),t}(x,y,\theta)}{\hat{f}_{\sigma(n),t}(x,y,\theta)} & \text{if} \begin{cases} \hat{f}_{\sigma(n),t}(x,y,\theta) \geq m_n \\ |(x,y')'| \leq \alpha_n \\ |\hat{f}'_{\sigma(n),t}(x,y,\theta)| \leq c_n \hat{f}_{\sigma(n),t}(x,y,\theta) \end{cases} \\ 0 & \text{otherwise} \end{cases}$$

where $c_n \to \infty, \alpha_n \to \infty, \sigma(n) \to 0, m_n \to 0.$ These conditions on the asymptotic behaviour of the smoothing parameter $\sigma(n)$ and the trimming parameters c_n, α_n, and m_n are used to show that our score estimator $\hat{\Delta}_n$, given in (21), is consistent. The trimming parameters serve to omit extreme outlying observations that would distort the behaviour of the estimate $\hat{\Delta}_n$.

At this point, a problem arises that is common in nonparametric estimation. The theory only describes the limiting behaviour of the smoothing and trimming parameters. In practical applications, with a fixed sample size, knowing this theoretical limiting behaviour provides little assistance in selecting the values to be used. An extensive literature exists regarding the selection of smoothing parameters in density estimation problems (see, for example, Marron (1987) for a survey), but the applicability of this literature to the case at hand has not been much investigated, nor has the question of trimming parameter selection.

The question of smoothing and trimming parameter selection in the adaptive estimation of (non-cointegrating) linear regression models was addressed in a Monte Carlo simulation study by Hsieh and Manski (1987), which extends a similar study reported by Manski (1984). Hsieh and Manski used sample sizes of 25 and 50, and considered six possible distributions for the errors (normal, variance contaminated mixture of normals, t, bimodal mixture of normals, beta, and log-normal). They set the standard deviation of the errors equal to unity. They found that the adaptive estimator's performance was fairly insensitive to the selection of trimming parameters (although being more sensitive to mild overtrimming than to mild undertrimming). They found that good values of c_n, α_n, and m_n for n=50 were 8,8, and exp(-32), respectively. Regarding the smoothing parameter $\sigma(n)$, they found that the estimator was quite sensitive to its selection. Depending on the true distribution of the errors, it was found that from a set of preselected possible values of $\sigma(n)$, the best value was anywhere from 0.1 to 0.5 for n=50. It was also found that estimator performance improved if a data-based bootstrap method was used to select $\sigma(n)$. Hsieh and Manski (1987) concluded by recommending the use of such a method in empirical applications and strongly recommending against using preselected values for $\sigma(n)$.

We now turn to the use of the above error density score estimator for the consistent estimation of the score function and information matrix of the entire model. Define

$$(21)\qquad \hat{\Delta}_n(\theta) = -\sum_{t=1}^{n} U_{nt}(\theta)\hat{\psi}(\varepsilon_t(\theta), v_t)$$

$$(22)\qquad \Delta_n(\theta) = -\sum_{t=1}^{n} U_{nt}(\theta)\psi(\varepsilon_t(\theta), v_t).$$

From (16), we have

$$(23)\qquad \frac{S_n(\theta)}{\lambda^2} = \sum_{t=1}^{n} \delta_n H_{t-1}(\theta) H_{t-1}(\theta)'\,\delta_n.$$

With the foregoing notation, we can now derive an adaptive estimator for our model. We begin by introducing the following three conditions and a result of Jeganathan (1988):

<u>Condition 5.1</u> *(condition (28) in Jeganathan (1988, p. 35)):*

$$0 < \int_{-\infty}^{\infty} \left[\frac{f_x(x,y)}{f(x,y)}\right]^2 f(x,y)dx < \infty.$$

<u>Condition 5.2</u> *(condition (B.2) in Jeganathan (1988, pp. 38-39)):*

We must verify that there is a suitable sequence $\{\delta_n\}$ *of normalizing matrices such that for every bounded* $\{h_n\}$ *(where* $\theta_n = \theta + \delta_n h_n$*) we have:*

$$(24) \quad \sum_{t=1}^{n}\left[\left(g_{t-1}(\theta_n)-g_{t-1}(\theta)\right)-h_n' U_{nt}(\theta)\right]^2 = o_p(1) \ \ in\ P_{\theta,n}$$

for suitable (p+q+m)-vectors $U_{n1},\ldots,U_{nn}$ *such that*

$$(25) \quad \sum_{t=1}^{n}\left|h_n' U_{nt}(\theta)\right|^2 = O_p(1) \ \ in\ P_{\theta,n}, \ \ and$$

$$(26) \quad \sup_{t\in\{1,\ldots,n\}}\left|h_n' U_{nt}(\theta)\right|^2 = o_p(1) \ \ in\ P_{\theta,n}.$$

<u>Condition 5.3</u> *(condition (B.3) in Jeganathan (1988, pp. 44-45)):*

Verify that there are s-vectors $V_{nt}(\theta), t=1,\ldots,n,$ *and non-random (p+q+m-s)-vectors* $R_{nt}(\theta), t=1,\ldots,n,$ *such that for every bounded* $\{h_n\}$ *and for every u, we have*

$$(27) \quad \sum_{t=1}^{n}\left|u' U_{nt}(\theta_n) - u'\begin{bmatrix} V_{nt}(\theta) \\ R_{nt}(\theta)\end{bmatrix}\right|^2 = o_p(1) \ \ in\ P_{\theta,n}$$

and, for some $\delta \in [0,1)$,

$$(28) \quad \max_{t\in\{1,\ldots,n\}} n\left|V_{nt}(\theta)\right|^2 = O_p(n^{\delta}) \ \ in\ P_{\theta,n}.$$

We now state the following proposition (based on Proposition 15 of Jeganathan (1988, pp. 46-50)):

<u>Proposition 5.5</u>: *Assume that Conditions 5.1, 5.2, and 5.3 hold and that f(z,v) is symmetric about zero in z for given v. Further assume that* $c_n \to \infty$, $\alpha_n \to \infty$, $m_n \to 0$,

$\sigma(n) \to 0$, $\sigma(n)c_n \to 0$, *and* $n^{-(1-\delta)}\alpha_n \sigma(n)^{-(5+m)} \to 0$, *with* δ *as in (28). Furthermore, assume that the sequences* $\{P_{\theta,n}\}$ *and* $\{P_{\theta_n,n}\}$ *are contiguous for every bounded* $\{h_n\}$. *Then, for every bounded* $\{h_n\}$, *the following hold:*

(29) $\quad \hat{\Delta}_n(\theta_n) - \Delta_n(\theta_n) = o_p(1)$ *in* $P_{\theta,n}$,

(30) $\quad \hat{\Delta}_n(\theta_n) = \hat{\Delta}_n(\theta) - S_n(\theta)h_n + o_p(1)$ *in* $P_{\theta,n}$, *and*

(31) $\quad \dfrac{S_n(\theta_n)}{\lambda^2} = \dfrac{S_n(\theta)}{\lambda^2} + o_p(1)$ *in* $P_{\theta,n}$.

The condition $n^{-(1-\delta)}\alpha_n \sigma(n)^{-(5+m)} \to 0$ is not exactly the same as the analogous condition given in Jeganathan's (1988) statement of the proposition. How this changes the proof of the proposition is described on p. 74 of Jeganathan (1988).

Using the results of Proposition 5.4, we can prove the following theorem, in which an expression for an adaptive estimator of the parameters of our model is derived.

<u>Theorem 5.5</u>: By setting $U_{nt}(\theta) = \delta_n H_{t-1}(\theta)$ *in equations (21) and (22), the following estimator is adaptive for our model:*

(32) $\quad \tilde{\theta}_n = \theta_n^{**} + \delta_n \left(\hat{I}_n \dfrac{S_n(\theta_n^{**})}{\lambda^2} \right)^{-1} \hat{\Delta}_n(\theta_n^{**})$

where $\hat{I}_n = \lambda^2 + o_p(1)$ *in* $P_{\theta,n}$. *(Computation of* $\hat{I}_n$ *will be described below.)*

In other words,

$$(33) \qquad \delta_n^{-1}(\tilde{\theta}_n - \hat{\theta}_n) = o_p(1) \ \ in\ P_{\theta,n}$$

where $\hat{\theta}_n$ *is as in (18).*

Kreiss (1987b, p. 123) shows that $\hat{I}_n$ as defined in equation (34) below is a consistent estimator of λ^2 (the proof uses the fact that $\{P_{\theta,n}\}$ and $\{P_{\theta_n,n}\}$ are contiguous):

$$(34) \qquad \hat{I}_n = \frac{1}{n}\sum_{t=1}^{n} \hat{\psi}_{n,t}^2\left(z_t(\theta_n^{**}), v_t, \theta_n^{**}\right).$$

We now derive the asymptotic distribution of our adaptive estimator $\tilde{\theta}_n$. Since $\tilde{\theta}_n$ is an AC estimator, we can use the definition of the latter to get:

$$\delta_n^{-1}(\tilde{\theta}_n - \theta) = S_n^{-1}(\theta)W_n(\theta) + o_p(1) \ \ in\ P_{\theta,n}.$$

Since we have (from equation (14) and the ensuing discussion)

$$L\left(S_n^{-1}(\theta)W_n(\theta)\middle|P_{\theta,n}\right) \Rightarrow L\left(\left(\lambda^2 J(\theta)\right)^{-1/2} N(0,I)\right),$$

it follows that

$$(35) \qquad L\left(\delta_n^{-1}(\tilde{\theta}_n - \theta)\middle|P_{\theta,n}\right) \Rightarrow L\left(MN\left(0, \frac{J^{-1}(\theta)}{\lambda^2}\right)\right).$$

We therefore have

$$(36)\qquad L\left(\sqrt{n}(\tilde{\eta}_n - \eta)\middle|P_{\theta,n}\right) \Rightarrow L\left(N\left(0, \frac{J_{ZZ}^{-1}(\theta)}{\lambda^2}\right)\right)$$

and

$$(37)\qquad L\left(n(\tilde{B}_n - B)\middle|P_{\theta,n}\right) \Rightarrow L\left(MN\left(0, \frac{J_{\Gamma\Gamma}^{-1}(\theta)}{\lambda^2}\right)\right).$$

The block diagonality of $J(\theta)$ is important because it will often be the case in applications that B is the parameter of interest and η is a nuisance parameter. Block diagonality implies that not knowing η implies no loss in asymptotic efficiency in estimating B vis-a-vis the case where η is known. In other words, if we regard the nuisance parameters of the model as being the infinite-dimensional nuisance parameter $p(\varepsilon, \nu)$ *and* the finite vector η, then $\tilde{B}_n$ is an adaptive estimator of B.

Another consequence of this block diagonality is that, if we are only interested in estimating B, we need not construct the entire estimate $\tilde{\theta}_n$ given in (32). We need only estimate the lower right-hand $m \times m$ submatrix of $S_n(\theta_n^{**})$ and the final m elements of the vector $\hat{\Delta}_n(\theta_n^{**})$, since the off-block-diagonal elements of $S_n(\theta_n^{**})$ converge in probability to zeros. In other words, we can compute the following adaptive estimate of B:

$$(38) \quad \widetilde{B}_n^* = B_n^{**} + \frac{1}{n}\left(\hat{I}_n \sum_{t=1}^{n} \frac{\Gamma_{t-1}(\theta_n^{**})\Gamma_{t-1}(\theta_n^{**})'}{n^2}\right)^{-1}\left(-\frac{1}{n}\sum_{t=1}^{n}\Gamma_{t-1}(\theta_n^{**})\hat{\psi}_{n,t}(z_t(\theta_n^{**}), v_t, \theta_n^{**})\right),$$

such that $n\left(\widetilde{B}_n^* - \widetilde{B}_n\right) = o_p(1)$ in $P_{\theta,n}$.

We now give an example to illustrate the gain in asymptotic efficiency that can be obtained from employing the adaptive estimator rather than the Gaussian pseudo-MLE. For the illustration, we assume that $p(\varepsilon, v)$ follows a multivariate *t*-distribution. Suppose that the deviations of the system from its cointegrating relationship are iid, so that $u_t = \varepsilon_t$. Further suppose that $\sigma_{21} = 0$ and that $p(\varepsilon, v)$ is the density of an $(m+1)$-dimensional *t*-distribution with $\tau > 2$ degrees of freedom. The asymptotic covariance matrix of the scaled and centred Gaussian MLE (OLS in this example) is $k\omega_{11}\left(\int_0^1 M_2 M_2'\right)^{-1}$, where $k = \tau/(\tau-2)$, while the asymptotic covariance matrix of the MLE, and therefore of $\widetilde{B}_n$, is $\lambda^{-2}\int_0^1 M_2 M_2'$, where $\lambda^{-2} = \omega_{11}\left(\frac{\tau+m+3}{\tau+m+1}\right)$.

One way to evaluate the relative efficiency of the two estimators is to take the ratio of the determinants of their asymptotic covariance matrices. In this example, the problem then reduces to taking the ratio between λ^{-2} and $k\omega_{11}$. We then have (see Mitchell, 1989):

$$\begin{aligned}\lambda^{-2}/k\omega_{11} &= \left(\frac{\tau+m+3}{\tau+m+1}\right)\cdot\left(\frac{\tau-2}{\tau}\right)\\ &= (1-2/\tau)\cdot(1+2/(\tau+m+1)).\end{aligned}$$

This quantity is decreasing in the number of regressors so that the more variables are included in the model, the greater the loss of asymptotic efficiency through the use of the OLS estimator.

In finite samples, however, there is good reason to expect that the performance of the adaptive estimator will worsen as m increases. This is because, in deriving our adaptive estimator, we employ a kernel estimator for a density of dimension $m+1$. From a computational standpoint, this kernel estimator can perform poorly when m is large. For sample sizes typical in econometrics, m need only equal three or four for our kernel estimator to give inaccurate results.

One way to alleviate this dimensionality problem is to take advantage of our elliptical symmetry assumption. One property of elliptically symmetric densities is that they can be expressed as a function of a scalar random variable, where the latter is a quadratic term in the underlying vector-valued random variable (see Fang, Kotz, and Ng (1990, p. 46)). In our case, we have from above that:

$$\begin{aligned} p(\varepsilon, v) &= \left|\det \Omega\right|^{-1/2} f^{*}\left(\left|\Omega^{-1/2}\begin{pmatrix}\varepsilon \\ v\end{pmatrix}\right|\right) \\ &= \left|\det \Omega\right|^{-1/2} f^{*}\left(z^{2}\ell_{11}^{-2} + v' \, \Omega_{22}^{-1} v\right). \end{aligned}$$

We could therefore proceed by using a normal kernel estimator such as the one used in the multivariate case to get an estimate $\hat{f}^{*}$ of f^{*}. We could then estimate $\hat{\psi}$ as above, using the derivative $2z\ell_{11}^{-2}\hat{f}^{*\prime}$. Proposition 1 would still apply, with the condition $n^{-(1-\delta)}\alpha_n\sigma(n)^{-(5+m)} \to 0$ being replaced by $n^{-(1-\delta)}\alpha_n\sigma(n)^{-5} \to 0$. In doing the computations, we would replace Ω with a consistent preliminary estimate $\hat{\Omega}$ throughout.

VI. INCLUSION OF A CONSTANT TERM

One limitation of the model given in equation (1) is that the regression lacks an intercept. In this section, we generalize the analysis through the addition of an intercept. It is shown that the asymptotic covariance matrix for $\widetilde{B}_n$, the adaptive estimate of the slope parameters, differs in the present case from the covariance matrix in the no-intercept case by a positive-definite matrix. It is shown that if the intercept is treated as a nuisance parameter, then the slope vector B *cannot* be adaptively estimated. We can estimate B more efficiently if the intercept is known.

We modify equation (1) as follows:

$$(1') \qquad Y_t = B_0 + BX_t + u_t,$$

with our other assumptions remaining the same. Carrying through the analysis of Section II, we arrive at the following modification of equation (7):

$$(7') \qquad \begin{aligned} \varepsilon_t(\theta_n) - \varepsilon_t(\theta) &= (\eta - \eta_n)' Z_{t-1}(\eta_n, \eta, B_0, B) + (B_0 - B_{0n})\bar{\gamma}_n \\ &\quad + (B - B_n)' \Gamma^*_{t-1}(\eta_n, B_0, B), \end{aligned}$$

where $\bar{\gamma}_n = \left(\sum_{j=0}^{\infty} \gamma_j(B_0, B, \eta_n)\right)\left(1 - \sum_{k=1}^{p} a_k^n\right)$, $\theta_n = (\eta_n', B_{0n}, B_n')'$, and $B_{0n} - B_0 = \frac{1}{\sqrt{n}} h_{B_0 n}$.

It follows that

$$
(9') \quad \begin{aligned} d_t(\theta_n, \theta) &= (\eta_n - \eta)' Z_{t-1} + (B_{0n} - B_0)\bar{\gamma}_n + (B_n - B)' \Gamma^*_{t-1} \\ &= (\theta_n - \theta)' H^*_{t-1} = h_n' \delta_n H^*_{t-1}, \end{aligned}
$$

where $H^*_{t-1} = (Z_{t-1}', \bar{\gamma}_n, \Gamma^*_{t-1}')'$, $\delta_n = diag(n^{-1/2} I_{p+q+1}, n^{-1} I_m)$, and $h_n = (h_{\eta n}', h_{B_0 n}, h_{Bn}')'$.

We then arrive at the following modification of Theorem 3.1:

<u>*Proposition 6.1*</u>*: In the model with intercept, the following approximation obtains:*

$$
(11') \quad \begin{aligned} \Lambda_n(\theta_n, \theta) = &-\sum_{t=1}^{n} h_n' \delta_n H_{t-1}(\theta) \psi(\varepsilon_t, v_t) \\ &- \frac{\lambda^2}{2} h_n' \delta_n \sum_{t=1}^{n} H_{t-1}(\theta) H_{t-1}(\theta)' \delta_n h_n + o_p(1) \ \ in\ P_{\theta,n}, \end{aligned}
$$

where $H_{t-1}(\theta) = (Z_{t-1}(\eta, \eta, B_0 B)', \bar{\gamma}, \Gamma_{t-1}(\eta, B_0, B)')'$ *and* $\bar{\gamma} = \left(\sum_{j=0}^{\infty} \gamma_j(B_0, B, \eta)\right)\left(1 - \sum_{k=1}^{p} a_k\right)$.

As above, we can show that

$$
(87') \quad \delta_n \sum_{t=1}^{n} H_{t-1}(\theta) H_{t-1}(\theta)' \delta_n \Rightarrow \begin{bmatrix} J_{ZZ}(\theta) & 0 & 0 \\ 0 & \bar{\gamma}^2 & \bar{\gamma} c_p \gamma(1) \int_0^1 M_2' \\ 0 & \bar{\gamma} c_p \gamma(1) \int_0^1 M_2 & J_{\Gamma\Gamma}(\theta) \end{bmatrix} \equiv J_1(\theta).
$$

Our LAMN limit theory holds since

$$L\left(\sum_{t=1}^{n}\delta_n H_{t-1}\psi(\varepsilon_t,v_t)\middle|P_{\theta,n}\right)\Rightarrow L\left(MN\left(0,\lambda^2 J_1(\theta)\right)\right).$$

The formula for our adaptive estimator $\widetilde{\theta}_n$ is still given by (32), using our redefinitions in (9') and (11') of $h_n'\delta_n H_{t-1}$. The asymptotic distributions of our estimators are as follows:

$$(35')\quad L\left(\delta_n^{-1}(\widetilde{\theta}_n-\theta)\middle|P_{\theta,n}\right)\Rightarrow L\left(MN\left(0,\frac{J_1(\theta)^{-1}}{\lambda^2}\right)\right),$$

$$(36')\quad L\left(\sqrt{n}(\widetilde{\eta}_n-\eta)\middle|P_{\theta,n}\right)\Rightarrow L\left(N\left(0,\frac{J_{ZZ}(\theta)^{-1}}{\lambda^2}\right)\right),$$

$$(39)\quad L\left(\sqrt{n}(\widetilde{B}_{0n}-B_0)\middle|P_{\theta,n}\right)\Rightarrow L\left(MN\left(0,\bar{\gamma}^{-2}\left[1+\int_0^1 M_2'\left(\int_0^1 M_2M_2'-\int_0^1 M_2\int_0^1 M_2'\right)^{-1}\int_0^1 M_2\right]\right)\right),$$

and

$$(37')\quad L\left(n(\widetilde{B}_n-B)\middle|P_{\theta,n}\right)\Rightarrow L\left(MN\left(0,\frac{1}{\lambda^2 c_p^2\gamma(1)^2}\left(\int_0^1 M_2M_2'-\int_0^1 M_2\int_0^1 M_2'\right)^{-1}\right)\right).$$

As mentioned above, our estimate $\widetilde{B}_n$ is not adaptive if B_0 is considered to be a nuisance parameter. This can be seen in two ways. First, the component of the information matrix $J_1(\theta)$ associated with the vector $(B_0,B')'$ is not block diagonal.

Second, the covariance matrix of $\widetilde{B}_n$ in (37') differs from that of $\widetilde{B}_n$ in (37),

$\lambda^{-2}c_p^{-2}\gamma(1)^{-2}\left(\int_0^1 M_2M_2'\right)^{-1}$, by a positive definite matrix (with probability one).

VII. EXTENSIONS AND GENERALIZATIONS

There are two immediate directions in which the results above could be extended. In the first place, the first differences in the regressors are assumed to follow an iid process. It would be desirable to extend the present analysis to allow for time dependence in these first differences. Maintaining a fully parametric specification, we could allow the first differences in the regressors and the innovations to the cointegrating regression to follow a joint vector ARMA process. Carrying out such an analysis would require results on the adaptive estimation of vector ARMA processes (an extension of Kreiss (1987b) to the multivariate case). Alternatively, and more interestingly, we could explore the possibility of treating the time dependence and endogeneity nonparametrically, as in Phillips and Hansen (1990).

A second extension would involve relaxing our assumption of an elliptically symmetric density function for the errors. This assumption (or that of symmetry in the univariate case) is used in the proofs of Bickel (1982), Kreiss (1987b), and Jeganathan (1988). However, Kreiss (1987a) uses an alternative method of proof which allows for an asymmetric density function in the univariate case. It would be desirable to investigate the application of this method to the present analysis, allowing for a class of density functions more general than the elliptically symmetric.

APPENDIX

Proof of Theorem 3.1: We begin by stating Conditions A.1-A.5, Proposition A.6, and Lemmas A.7-A.9. These results will be used in Lemmas A.10 and A.11, which together prove the theorem. Conditions A.1-A.5 are specializations to our model of Conditions (A.1)-(A.5) of Jeganathan (1994).

Condition A.1: *The conditional density of* ε *given* v, $\dfrac{p(\varepsilon,v)}{e(v)}$, *is absolutely continuous in* ε, *where* $e(v)$ *is the marginal Lebesgue density of* v.

Condition A.2: *The derivative* $\dfrac{\partial p(\varepsilon,v)}{\partial\varepsilon}\Big/ e(v)$ *exists.*

Condition A.3: *If* $\{\theta_n\}\subseteq\Theta$ *is a sequence such that*

$$(40)\qquad \sum_{t=1}^{n} E\left[\left|d_t'(\theta_n,\theta)\psi(\varepsilon_t,v_t)\right|^2 \middle| A_{t-1}\right] = O_p(1) \quad in\ P_{\theta,n},$$

then the quantities

$$(41)\qquad \sum_{t=1}^{n}\int\int_0^1 \left\|d_t'(\theta_n,\theta)\left[\psi^*\left(\varepsilon-\kappa d_t(\theta_n,\theta),v_t\right)-\psi^*\left(\varepsilon,v_t\right)\right]\right\|^2 d\varepsilon d\kappa = o_p(1) \quad in\ P_{\theta,n},$$

where $\psi^*(\varepsilon,v)=\dfrac{\left(\dfrac{\partial p(\varepsilon,v)}{\partial\varepsilon}\Big/ e(v)\right)}{\sqrt{p(\varepsilon,v)/e(v)}}$,

and, $\forall\omega>0$,

$$\sum_{t=1}^{n} E\left[\left|d_t'(\theta_n,\theta)\psi(\varepsilon_t,v_t)\right|^2 I\left(\left|d_t'(\theta_n,\theta)\psi(\varepsilon_t,v_t)\right| > \omega\right)\middle|A_{t-1}\right] = o_p(1) \quad in\, P_{\theta,n}. \tag{42}$$

<u>Condition A.4</u>: $E\left[\psi(\varepsilon_t,v_t)\middle|A_{t-1}\right]=0,\ t\geq 1.$

<u>Condition A.5</u>: $f_0\left(\underline{Y}_0,\theta_n\right)-f_0\left(\underline{Y}_0,\theta\right)=o_p(1)\ \ in\ P_{\theta,n}\ as\ \theta_n\to\theta.$

<u>Proposition A.6</u> (Theorem 11 in Jeganathan (1994)):

Assume that Conditions A.1-A.5 hold. Then, for every $\{\theta_n\}$ *such that (40) holds, we have*

$$\begin{aligned}
&\Lambda_n(\theta_n,\theta)-\left\{-\sum_{t=1}^{n}\left(g_{t-1}(\theta_n)-g_{t-1}(\theta)\right)'\psi(\varepsilon_t,v_t)\right.\\
&\qquad\left.-\frac{1}{2}\sum_{t=1}^{n}E\left[\left|\left(g_{t-1}(\theta_n)-g_{t-1}(\theta)\right)'\psi(\varepsilon_t,v_t)\right|^2\middle|A_{t-1}\right]\right\}\\
&\qquad=o_p(1)\ \ in\,P_{\theta,n}.
\end{aligned}\tag{43}$$

<u>Lemma A.7</u> (Lemma 19 in Jeganathan (1994)):

Let W(y) be Lebesgue measurable such that $\int\left|W(y)\right|^2 dy<\infty$. *Then*

$$\int\left|W\left(\frac{y+\omega}{\delta}\right)-W(y)\right|^2 dy\to 0$$

as $\omega\to 0$ *and* $\delta\to 1$.

<u>Lemma A.8</u> (Lemma 24 in Jeganathan (1994)):

For each $n \geq 1$, *let* $(\xi_{n1}, \ldots, \xi_{nn})$ *be an array of random variables and let* $(\beta_{n1}, \ldots, \beta_{nn})$ *be an array of* $\sigma-$*fields such that* $\beta_{n1} \subseteq \ldots \subseteq \beta_{nn}$ *and* ξ_{nt} *is* $\beta_{nt}-$*measurable. Furthermore, let* β_{n0} *be the trivial* $\sigma-$*field. Assume that*

$$(44) \quad \sum_{t=1}^{n} E\left[\xi_{nt}^2 \middle| \beta_{n,t-1}\right] = O_p(1)$$

and, $\forall \omega > 0$,

$$(45) \quad \sum_{t=1}^{n} E\left[\xi_{nt}^2 \cdot I\left(|\xi_{nt}| > \omega\right) \middle| \beta_{n,t-1}\right] = o_p(1).$$

Then

$$(46) \quad \sum_{t=1}^{n} E\left[\xi_{nt}^2 \middle| \beta_{n,t-1}\right] - \sum_{t=1}^{n} \xi_{nt}^2 = o_p(1).$$

<u>*Lemma A.9*</u> *(equation (2.32) on p. 46 of Hall and Heyde (1980))*:

Let ξ_{nt} *and* β_{nt}, $t=1,\ldots,n$, *be as in Lemma A.8. Then, for any constants* $\tau > 0$ *and* $\omega > 0$,

$$P\left(\max_{t \in \{1,\ldots,n\}} |\xi_{nt}| > \tau\right) \leq \omega + P\left(\sum_{t=1}^{n} E\left[\xi_{nt}^2 \cdot I\left(|\xi_{nt}| > \tau\right) \middle| \beta_{n,t-1}\right] > \tau^2 \omega\right).$$

We shall complete the proof of Theorem 3.1 in two steps. We first show, in Lemma A.10, that if Conditions A.1-A.5 are satisfied, then (11) is implied by Proposition A.6.

We then show, in Lemma A.11, that Conditions A.1-A.5 are indeed satisfied for our model.

Lemma A.10: *Equation (43) implies equation (11).*

Proof: Substituting (9) into equation (43) above, we have

$$\Lambda_n(\theta_n,\theta) = -\sum_{t=1}^{n}[(\theta_n-\theta)' H_{t-1}^{*}\psi(\varepsilon_t,v_t)] - \frac{1}{2}\sum_{t=1}^{n} E[|(\theta_n-\theta)' H_{t-1}^{*}\psi(\varepsilon_t,v_t)|^2 \, |A_{t-1}] + o_p(1) \ \ in\ P_{\theta,n}. \tag{47}$$

We now show that the right-hand side of (47) can be rewritten as

$$-\sum_{t=1}^{n}[(\theta_n-\theta)' H_{t-1}\psi(\varepsilon_t,v_t)] - \frac{1}{2}\sum_{t=1}^{n} E[|(\theta_n-\theta)' H_{t-1}\psi(\varepsilon_t,v_t)|^2 \, |A_{t-1}] + o_p(1) \ \ in\ P_{\theta,n}. \tag{48}$$

Noting that $H_{t-1}^{*} = H_{t-1} + \begin{bmatrix} 0 \\ v_t \end{bmatrix}$, we have

$$\begin{aligned}
&\sum_{t=1}^{n}\left[(\theta_n-\theta)' H_{t-1}^{*}\psi(\varepsilon_t,v_t)\right] \\
&= \sum_{t=1}^{n}\left[(\theta_n-\theta)'\left[H_{t-1} + \begin{bmatrix} 0 \\ v_t \end{bmatrix}\right]\psi(\varepsilon_t,v_t)\right] \\
&= \sum_{t=1}^{n}\left[(\theta_n-\theta)' H_{t-1}\psi_t + (\theta_n-\theta)'\begin{bmatrix} 0 \\ v_t \end{bmatrix}\psi(\varepsilon_t,v_t)\right] \\
&= \sum_{t=1}^{n}\left[(\theta_n-\theta)' H_{t-1}\psi(\varepsilon_t,v_t)\right] + o_p(1) \ \ in\ P_{\theta,n},
\end{aligned}$$

since

$$\sum_{t=1}^{n}\left[(\theta_n-\theta)'\begin{bmatrix}0\\ v_t\end{bmatrix}\psi(\varepsilon_t,v_t)\right]=\sum_{t=1}^{n}\left[h_n'\,\delta_n\begin{bmatrix}0\\ v_t\psi(\varepsilon_t,v_t)\end{bmatrix}\right]$$
$$=h_{Bn}'\sum_{t=1}^{n}\frac{v_t\psi(\varepsilon_t,v_t)}{n}=o_p(1)\ \ in\ P_{\theta,n},$$

the final equality holding because $\{v_t\psi(\varepsilon_t,v_t)\}$ is an iid sequence, because

$$E\left[\left|v_1\psi(\varepsilon_1,v_1)\right|\right]\le\left(E\left[v_1^2\right]\right)^{1/2}\left(E\left[\psi(\varepsilon_1,v_1)^2\right]\right)^{1/2}<\infty$$

(by the Cauchy-Schwartz inequality and our assumptions), and because, as shown by Jeganathan (1988, p. 69), $E\left[v_1\psi(\varepsilon_1,v_1)\right]=0.$

We also have

$$\frac{1}{2}\sum_{t=1}^{n}E\left[\left|(\theta_n-\theta)'\,H_{t-1}^{*}\psi(\varepsilon_t,v_t)\right|^2\middle|A_{t-1}\right]$$
$$(49)\qquad =\frac{1}{2}\sum_{t=1}^{n}E\left[(\theta_n-\theta)'\left(H_{t-1}+\begin{bmatrix}0\\ v_t\end{bmatrix}\right)\left(H_{t-1}+\begin{bmatrix}0\\ v_t\end{bmatrix}\right)'(\theta_n-\theta)\psi^2(\varepsilon_t,v_t)\middle|A_{t-1}\right]$$
$$=\frac{1}{2}\sum_{t=1}^{n}E\left[(\theta_n-\theta)'\,H_{t-1}H_{t-1}'\,(\theta_n-\theta)\psi^2(\varepsilon_t,v_t)\middle|A_{t-1}\right]+o_p(1)\ \ in\ P_{\theta,n}.$$

The $o_p(1)$ result in the final line of (49) holds due to the final line of (50) below and to the final line of (57) below. We begin the verification of (49) by noting that

$$\frac{1}{2}\sum_{t=1}^{n} E\left[(\theta_n-\theta)'\begin{bmatrix}0\\ v_t\end{bmatrix}H_{t-1}'(\theta_n-\theta)\psi^2(\varepsilon_t,v_t)\middle|A_{t-1}\right]$$

$$=\frac{1}{2}\sum_{t=1}^{n} E\left[(\theta_n-\theta)'\begin{bmatrix}0\\ v_t\end{bmatrix}\begin{bmatrix}Z_{t-1}' & \Gamma_{t-1}'\end{bmatrix}(\theta_n-\theta)\psi^2(\varepsilon_t,v_t)\middle|A_{t-1}\right]$$

$$=\frac{1}{2}h_n'\sum_{t=1}^{n} E\left[\begin{bmatrix}\frac{1}{\sqrt{n}}I & 0\\ 0 & \frac{1}{n}I\end{bmatrix}\begin{bmatrix}0\\ v_t\end{bmatrix}\begin{bmatrix}Z_{t-1}' & \Gamma_{t-1}'\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{n}}I & 0\\ 0 & \frac{1}{n}I\end{bmatrix}\psi^2(\varepsilon_t,v_t)\middle|A_{t-1}\right]h_n$$

$$=\frac{1}{2}h_n'\sum_{t=1}^{n} E\left[\begin{bmatrix}0 & 0\\ \frac{v_tZ_{t-1}'}{n^{3/2}} & \frac{v_t\Gamma_{t-1}'}{n^2}\end{bmatrix}\psi^2(\varepsilon_t,v_t)\middle|A_{t-1}\right]h_n$$

$$=\frac{1}{2}h_n'\begin{bmatrix}0 & 0\\ \sum_{t=1}^{n} E\left[\frac{v_tZ_{t-1}'\psi^2(\varepsilon_t,v_t)}{n^{3/2}}\middle|A_{t-1}\right] & \sum_{t=1}^{n} E\left[\frac{v_t\Gamma_{t-1}'\psi^2(\varepsilon_t,v_t)}{n^2}\middle|A_{t-1}\right]\end{bmatrix}h_n$$

$$=o_p(1)\ in\ P_{\theta,n}. \tag{50}$$

We need to show that the two non-zero blocks of the matrix in the second last line of equation (50) are $o_p(1)$ in $P_{\theta,n}$, i.e. that the following two conditions hold:

$$\sum_{t=1}^{n} E\left[\frac{v_tZ_{t-1}'\psi^2(\varepsilon_t,v_t)}{n^{3/2}}\middle|A_{t-1}\right]=o_p(1)\ in\ P_{\theta,n}, \tag{51}$$

$$\frac{1}{n^2}\sum_{t=1}^{n} E\left[v_t\Gamma_{t-1}'\psi^2(\varepsilon_t,v_t)\middle|A_{t-1}\right]=o_p(1)\ in\ P_{\theta,n}. \tag{52}$$

We begin with (51). Consider the $(i,\ell)^{th}$ element of the matrix on its left-hand side:

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}E\left[v_{it}Z_{\ell,t-1}\psi^{2}(\varepsilon_{t},v_{t})\middle|A_{t-1}\right]$$

$$=\begin{cases}\dfrac{1}{n^{3/2}}\displaystyle\sum_{t=1}^{n}E\left[v_{it}\left(\sum_{k=0}^{t-1}\gamma_{k}(B,\eta_{n})u_{t-k-\ell}\right)\psi^{2}(\varepsilon_{t},v_{t})\middle|A_{t-1}\right] & if\ 1\le\ell\le p\\ \dfrac{1}{n^{3/2}}\displaystyle\sum_{t=1}^{n}E\left[v_{it}\left(\sum_{k=0}^{t-1}\gamma_{k}(B,\eta_{n})\varepsilon_{t-k-\ell+p}\right)\psi^{2}(\varepsilon_{t},v_{t})\middle|A_{t-1}\right] & if\ p+1\le\ell\end{cases}$$

We shall prove the $o_p(1)$ result for the first case only, since an identical argument can be used for the second case.

We have

$$\begin{aligned}&\left|\frac{1}{n^{3/2}}\sum_{t=1}^{n}E\left[v_{it}Z_{\ell,t-1}\psi^{2}(\varepsilon_{t},v_{t})\middle|A_{t-1}\right]\right|\\&\le\frac{1}{n^{3/2}}\sum_{t=1}^{n}\left|E\left[v_{it}Z_{\ell,t-1}\psi^{2}(\varepsilon_{t},v_{t})\middle|A_{t-1}\right]\right|\\(53)\quad&\le\frac{1}{n^{3/2}}\sum_{t=1}^{n}E\left[\left|v_{it}Z_{\ell,t-1}\psi^{2}(\varepsilon_{t},v_{t})\right|\middle|A_{t-1}\right]\\&\le\left(\frac{1}{\sqrt{n}}\max_{t\in\{1,\ldots,n\}}\left|v_{it}Z_{\ell,t-1}\right|\right)\left(\frac{1}{n}\sum_{t=1}^{n}E\left[\psi^{2}(\varepsilon_{t},v_{t})\middle|A_{t-1}\right]\right)\\&=o_p(1)\cdot O_p(1)=o_p(1)\ \ in\ P_{\theta,n}.\end{aligned}$$

The $O_p(1)$ result follows since $E\left[\psi^{2}(\varepsilon_{t},v_{t})\middle|A_{t-1}\right]$ is an iid sequence with finite mean.

We must now show that

$$\frac{1}{\sqrt{n}}\max_{t\in\{1,\ldots,n\}}\left|\sum_{k=0}^{t-1}\gamma_{k}(B,\eta_{n})v_{it}u_{t-k-\ell}\right|=o_p(1)\ \ in\ P_{\theta,n}.$$

Now,

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} \max_{t\in\{1,\ldots,n\}} \left| \sum_{k=0}^{t-1} \gamma_k(B,\eta_n) v_{it} u_{t-k-\ell} \right| \\
&\le \frac{1}{\sqrt{n}} \max_{t\in\{1,\ldots,n\}} \sum_{k=0}^{t-1} \left| \gamma_k(B,\eta_n) v_{it} u_{t-k-\ell} \right| \\
&\le \sum_{k=0}^{t-1} \left| \gamma_k(B,\eta_n) \right| \frac{1}{\sqrt{n}} \max_{t\in\{1,\ldots,n\}} \left| v_{it} u_{t-k-\ell} \right| \\
&= o_p(1) \;\; in\, P_{\theta,n}.
\end{aligned}
$$

This final equality holds because

$$
\exists \Delta < \infty \; s.t. \sum_{k=0}^{t-1} \left| \gamma_k(B,\eta_n) \right| < \Delta \;\; \forall t \ge 1
$$

and because

$$
(54) \qquad \max_{t\in\{1,\ldots,n\}} \left| \frac{v_{it} u_{t-k-\ell}}{\sqrt{n}} \right| = o_p(1) \;\; in\; P_{\theta,n}.
$$

To verify (54), we apply Lemma A.9, setting $\xi_{nt} \equiv \dfrac{v_{it} u_{t-k-\ell}}{\sqrt{n}}$ and $F_{n,t} \equiv \sigma(\ldots(\varepsilon_t, v_t))$, so that we have

$$
\begin{aligned}
&P\left(\max_{t\in[1,\ldots,n]} \left| \frac{v_{it} u_{t-k-\ell}}{\sqrt{n}} \right| > \tau \right) \\
&\qquad \le \omega + P\left(\sum_{t=1}^{n} E\left[\frac{\left(v_{it} u_{t-k-\ell}\right)^2}{n} \cdot I\left(\left| \frac{v_{it} u_{t-k-\ell}}{\sqrt{n}} \right| > \tau \right) \middle| \beta_{n,t-1} \right] > \tau^2 \omega \right).
\end{aligned}
$$

To obtain our desired result, we must show that

$$
\frac{1}{n} \sum_{t=1}^{n} E\left[\left(v_{it} u_{t-k-\ell}\right)^2 \cdot I\left(\left| v_{it} u_{t-k-\ell} \right| > \tau\sqrt{n} \right) \middle| \beta_{n,t-1} \right] = o_p(1) \;\; in\; P_{\theta,n}.
$$

Since the summation is over a sequence of stationary and ergodic random variables with finite mean, it follows that

$$\frac{1}{n}\sum_{t=1}^{n} E\left[\left(v_{it}u_{t-k-\ell}\right)^2 \cdot I\left(\left|v_{it}u_{t-k-\ell}\right| > \tau\sqrt{n}\right)\middle|\beta_{n,t-1}\right]$$
$$- E\left[\left(v_{it}u_{t-k-\ell}\right)^2 \cdot I\left(\left|v_{it}u_{t-k-\ell}\right| > \tau\sqrt{n}\right)\right] = o_p(1) \ \ in\ P_{\theta,n}.$$

The result (54) follows from the fact that

$$E\left[\left(v_{it}u_{t-k-\ell}\right)^2 \cdot I\left(\left|v_{it}u_{t-k-\ell}\right| > \tau\sqrt{n}\right)\right] = o(1).$$

This verifies (51). To verify (52), we write the absolute value of the $(i,\ell)^{th}$ element of the matrix on its left-hand side as:

$$\left|\frac{1}{n^2}\sum_{t=1}^{n} E\left[v_{it}\Gamma_{\ell,t-1}\psi^2(\varepsilon_t, v_t)\middle|A_{t-1}\right]\right|$$
$$\leq \frac{1}{n^2}\sum_{t=1}^{n}\left|E\left[v_{it}\Gamma_{\ell,t-1}\psi^2(\varepsilon_t, v_t)\middle|A_{t-1}\right]\right|$$
$$\text{(55)} \qquad \leq \frac{1}{n^2}\sum_{t=1}^{n} E\left[\left|v_{it}\Gamma_{\ell,t-1}\psi^2(\varepsilon_t, v_t)\right|\middle|A_{t-1}\right]$$
$$\leq \left(\frac{1}{n}\max_{t\in\{1,\dots,n\}}\left|v_{it}\Gamma_{\ell,t-1}\right|\right)\left(\frac{1}{n}\sum_{t=1}^{n} E\left[\psi^2(\varepsilon_t, v_t)\middle|A_{t-1}\right]\right)$$
$$= o_p(1)\cdot O_p(1) = o_p(1) \ \ in\ P_{\theta,n}.$$

The first component of the second last line of (55) can be rewritten

$$\frac{1}{n}\max_{t\in\{1,\dots,n\}}\left|\sum_{j=0}^{t-1}\gamma_j(B,\eta_n)\left(X_{\ell,t-j}-\sum_{k=1}^{p}a_k^n X_{\ell,t-j-k}\right)v_{it}-v_{\ell t}v_{it}\right|$$

$$\le\frac{1}{n}\max_{t\in\{1,\dots,n\}}\left\{\left|\sum_{j=0}^{t-1}\gamma_j(B,\eta_n)\sum_{k=1}^{p}a_k^n X_{\ell,t-j-k}v_{it}\right|\right.$$

$$(56)\qquad \left.+\left|\sum_{j=0}^{t-1}\gamma_j(B,\eta_n)X_{\ell,t-j}v_{it}\right|+\left|v_{\ell t}v_{it}\right|\right\}$$

$$\le\frac{1}{n}\left\{\max_{t\in\{1,\dots,n\}}\sum_{j=0}^{t-1}\left|\gamma_j(B,\eta_n)\sum_{k=1}^{p}a_k^n X_{\ell,t-j-k}v_{it}\right|\right.$$

$$\left.+\max_{t\in\{1,\dots,n\}}\sum_{j=0}^{t-1}\left|\gamma_j(B,\eta_n)X_{\ell,t-j}v_{it}\right|+\max_{t\in\{1,\dots,n\}}\left|v_{\ell t}v_{it}\right|\right\}.$$

We now show that each of the three terms on the right-hand side of the final inequality in (56) is $o_p(1)$ in $P_{\theta,n}$. We begin with the third term, for which we can show in the same way as we showed (54) above that

$$(57)\qquad \max_{t\in\{1,\dots,n\}}\left|\frac{v_{\ell t}v_{it}}{n}\right|=o_p(1)\ \ in\ P_{\theta,n}.$$

We now consider the second term on the right-hand side of the final inequality in (56), showing that

$$(58)\qquad \max_{t\in\{1,\dots,n\}}\sum_{j=0}^{t-1}\left|\frac{\gamma_j(B,\eta_n)X_{\ell,t-j}v_{it}}{n}\right|=o_p(1)\ \ in\ P_{\theta,n}.$$

The left-hand side of (58) is less than or equal to

$$\sum_{j=0}^{t-1}\left|\gamma_j(B,\eta_n)\right|\max_{t\in\{1,\dots,n\}}\left|\frac{X_{\ell,t-j}v_{it}}{n}\right|=o_p(1)\ \ in\ P_{\theta,n}$$

since

$$\max_{t\in\{1,\ldots,n\}}\left|\frac{X_{\ell,t-j}v_{it}}{n}\right| \leq \left(\max_{t\in\{1,\ldots,n\}}\left|\frac{X_{\ell,t-j}}{\sqrt{n}}\right|\right)\left(\max_{t\in\{1,\ldots,n\}}\left|\frac{v_{it}}{\sqrt{n}}\right|\right)$$
$$= O_p(1)\cdot o_p(1) \ \ in\ P_{\theta,n},$$

where the $o_p(1)$ result can be verified in a manner analogous to the proof of (54) and the $O_p(1)$ result follows from (10.10) on p. 70 of Billingsley (1968).

We now consider the first term on the right-hand side of the final inequality in (56):

$$\max_{t\in\{1,\ldots,n\}}\sum_{j=0}^{t-1}\left|\frac{\gamma_j(B,\eta_n)\sum_{k=1}^{p}a_k^nX_{\ell,t-j-k}v_{it}}{n}\right|$$
$$\leq \sum_{k=1}^{p}a_k^n\left[\max_{t\in\{1,\ldots,n\}}\sum_{j=0}^{t-1}\left|\frac{\gamma_j(B,\eta_n)X_{\ell,t-j-k}v_{it}}{n}\right|\right]$$
$$= \sum_{k=1}^{p}a_k^n o_p(1) = o_p(1) \ \ in\ P_{\theta,n},$$

using (58).

We have shown (51), (52), and therefore the final equality of (50). We complete the verification of (49) by noting that:

$$\frac{1}{2}\sum_{t=1}^{n}E\left[(\theta_n-\theta)'\begin{pmatrix}0\\ v_t\end{pmatrix}(0 \quad v_t')(\theta_n-\theta)\psi^2(\varepsilon_t,v_t)\middle|A_{t-1}\right]$$

(59) $$=\frac{1}{2}h_n'\sum_{t=1}^{n}E\left[\begin{bmatrix}0 & 0\\ 0 & \frac{v_tv_t'}{n^2}\end{bmatrix}\psi^2(\varepsilon_t,v_t)\middle|A_{t-1}\right]h_n$$

$$=o_p(1) \;\; in\, P_{\theta,n},$$

because

(60) $$\frac{1}{n^2}\sum_{t=1}^{n}E\left[v_tv_t'\,\psi^2(\varepsilon_t,v_t)\middle|A_{t-1}\right]=o_p(1) \;\; in\, P_{\theta,n}.$$

The $(i,\ell)^{th}$ element of the left-hand side of (60) is

$$\frac{1}{n^2}\sum_{t=1}^{n}E\left[v_{it}v_{\ell t}\psi^2(\varepsilon_t,v_t)\middle|A_{t-1}\right]$$

$$\leq\left(\frac{\max_{t\in\{1,\dots,n\}}\left|v_{it}v_{\ell t}\right|}{n}\right)\left(\frac{1}{n}\sum_{t=1}^{n}E\left[\psi^2(\varepsilon_t,v_t)\middle|A_{t-1}\right]\right)$$

$$=o_p(1)\cdot O_p(1)=o_p(1) \;\; in\, P_{\theta,n},$$

by (53) and (57).

We have now established the equality of (48) and the right-hand side of (47). The negative of the second term in (48) is (using the notation that $W^2(v_t)\equiv E\left[\psi^2(\varepsilon_t,v_t)\middle|A_{t-1}\right]$):

$$
\begin{aligned}
&\frac{1}{2}\sum_{t=1}^{n} E\left[(\theta_n-\theta)' H_{t-1}H_{t-1}'(\theta_n-\theta)\psi^2(\varepsilon_t,v_t)\middle|A_{t-1}\right]\\
&=\frac{1}{2}\sum_{t=1}^{n}(\theta_n-\theta)' H_{t-1}H_{t-1}'(\theta_n-\theta)E\left[\psi^2(\varepsilon_t,v_t)\middle|A_{t-1}\right]\\
&=\frac{1}{2}\sum_{t=1}^{n}(\theta_n-\theta)' H_{t-1}H_{t-1}'(\theta_n-\theta)W^2(v_t)\\
&=\frac{\lambda^2}{2}\sum_{t=1}^{n}(\theta_n-\theta)' H_{t-1}H_{t-1}'(\theta_n-\theta)\\
&+\frac{1}{2}\sum_{t=1}^{n}(\theta_n-\theta)' H_{t-1}H_{t-1}'(\theta_n-\theta)\left(W^2(v_t)-\lambda^2\right)\\
&=\frac{\lambda^2}{2}\sum_{t=1}^{n}(\theta_n-\theta)' H_{t-1}H_{t-1}'(\theta_n-\theta)+o_p(1)\ \ in\ P_{\theta,n},
\end{aligned}
\tag{61}
$$

the final equality of (61) holding because

$$
\begin{aligned}
&\frac{1}{2}\sum_{t=1}^{n}(\theta_n-\theta)' H_{t-1}H_{t-1}'(\theta_n-\theta)\left(W^2(v_t)-\lambda^2\right)\\
&=\frac{1}{2}h_n'\sum_{t=1}^{n}\begin{bmatrix}\frac{Z_{t-1}Z_{t-1}'}{n} & \frac{Z_{t-1}\Gamma_{t-1}'}{n^{3/2}}\\ \frac{\Gamma_{t-1}Z_{t-1}'}{n^{3/2}} & \frac{\Gamma_{t-1}\Gamma_{t-1}'}{n^2}\end{bmatrix}h_n\left(W^2(v_t)-\lambda^2\right)\\
&=o_p(1)\ \ in\ P_{\theta,n}.
\end{aligned}
$$

To prove the final equality above, we must show that the following three results hold:

$$\frac{1}{n}\sum_{t=1}^{n}Z_{t-1}Z_{t-1}'\left(W^2(v_t)-\lambda^2\right)=o_p(1)\ \ in\ P_{\theta,n},\tag{62}$$

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}Z_{t-1}\Gamma_{t-1}'\left(W^2(v_t)-\lambda^2\right)=o_p(1)\ \ in\ P_{\theta,n},\ \text{and}\tag{63}$$

$$\frac{1}{n^2}\sum_{t=1}^{n}\Gamma_{t-1}\Gamma_{t-1}'\left(W^2(v_t)-\lambda^2\right)=o_p(1)\ \ in\ P_{\theta,n}.\tag{64}$$

To show (62), (63), and (64), note that

$$\frac{1}{n}\sum_{t=1}^{n}\left(W^2(v_t)-\lambda^2\right)=o_p(1)\ \ in\ P_{\theta,n}$$

because $W^2(v_t)$ is iid with mean λ^2. The results follow because it is easily shown that

$$(65)\qquad \max_{t\in\{1,\dots,n\}}\left|Z_{\ell,t-1}Z_{i,t-1}\right|=O_p(1)\ \ in\ P_{\theta,n}\ \ \forall\ell,i\in\{1,\dots,p+q\}$$

$$(66)\qquad \max_{t\in\{1,\dots,n\}}\left|\frac{Z_{\ell,t-1}\Gamma_{i,t-1}}{\sqrt{n}}\right|=O_p(1)\ \ in\ P_{\theta,n}\ \ \forall\ell\in\{1,\dots,p+q\},\ i\in\{1,\dots,m\}$$

$$(67)\qquad \max_{t\in\{1,\dots,n\}}\left|\frac{\Gamma_{\ell,t-1}\Gamma_{i,t-1}}{\sqrt{n}}\right|=O_p(1)\ \ in\ P_{\theta,n}\ \ \forall\ell,i\in\{1,\dots,m\}.$$

Results (11) and (48) are equivalent since $(\theta_n-\theta)=\delta_n h_n$. This completes the proof of Lemma A.10.

•

<u>Lemma A.11</u>: *Our model satisfies Conditions A.1-A.5 above.*

<u>Proof</u>: Conditions A.1 and A.2 are primitive conditions on $p(\varepsilon_t,v_t)$ that we assume to be satisfied.

We now verify condition A.3. We verify equations (40), (41), and (42) above, which can be rewritten as the following three equations, respectively:

(68) $$\sum_{t=1}^{n} E\left[\left|h_n{}'\,\delta_n H_{t-1}^{*}\psi(\varepsilon_t, v_t)\right|^2 \middle| A_{t-1}\right] = O_p(1) \ \ in\, P_{\theta,n},$$

(69) $$\sum_{t=1}^{n} \int_0^1 \int \left|h_n{}'\,\delta_n H_{t-1}^{*}\left[\psi^{*}\left(\varepsilon - \kappa h_n{}'\,\delta_n H_{t-1}^{*}, v_t\right) - \psi^{*}\left(\varepsilon, v_t\right)\right]\right\|^2 d\varepsilon d\kappa = o_p(1) in\, P_{\theta,n}.$$

and, $\forall \omega > 0$,

(70) $$\sum_{t=1}^{n} E\left[\left|h_n{}'\,\delta_n H_{t-1}^{*}\psi(\varepsilon_t, v_t)\right|^2 I(\left|h_n{}'\,\delta_n H_{t-1}^{*}\psi(\varepsilon_t, v_t)\right| > \omega) \middle| A_{t-1}\right] = o_p(1) \ \ in\, P_{\theta,n}.$$

We begin by verifying (68). It is sufficient to verify that:

(71) $$\sum_{t=1}^{n} E\left[\left|h_n{}'\,\delta_n H_{t-1}\psi(\varepsilon_t, v_t)\right|^2 \middle| A_{t-1}\right] = O_p(1) \ \ in\, P_{\theta,n}$$

and that

(72) $$\sum_{t=1}^{n} E\left[\left|h_n{}'\,\delta_n \begin{bmatrix} 0 \\ v_t \end{bmatrix} \psi(\varepsilon_t, v_t)\right|^2 \middle| A_{t-1}\right] = o_p(1) \ \ in\, P_{\theta,n}.$$

(72) has been verified above, so we need only verify (71). To do this, we proceed by showing that our model satisfies equations (44) and (45) of Lemma A.8, where we use the following notation:

$$\xi_{nt}^2 \equiv E\left[\left|h_n{}'\,\delta_n H_{t-1}\psi(\varepsilon_t, v_t)\right|^2 \middle| A_{t-1}\right],$$
$$\beta_{nt} \equiv \sigma((\varepsilon_j, v_j), j = 1, \ldots, t),$$

so that (44) and (45) become, respectively,

(73) $$\sum_{t=1}^{n} E\left[E\left[\left|h_n'\,\delta_n H_{t-1}\psi(\varepsilon_t,v_t)\right|^2 \middle| A_{t-1}\right] \middle| \beta_{n,t-1}\right] = O_p(1) \;\; in\, P_{\theta,n},$$

(74) $$\sum_{t=1}^{n} E\left[\begin{array}{l} E\left[\left|h_n'\,\delta_n H_{t-1}\psi(\varepsilon_t,v_t)\right|^2 \middle| A_{t-1}\right] \\ \cdot I\left(\sqrt{E\left[\left|h_n'\,\delta_n H_{t-1}\psi(\varepsilon_t,v_t)\right|^2 \middle| A_{t-1}\right]} > \omega\right) \middle| \beta_{n,t-1} \end{array}\right] = o_p(1) \;\; in\, P_{\theta,n} \quad \forall \omega > 0.$$

According to Lemma A.8, verification of (73) and (74) is sufficient to verify (71).

This is because if (73) and (74) hold, we get the result that

$$\begin{aligned} &\sum_{t=1}^{n} E\left[\left|h_n'\,\delta_n H_{t-1}\psi(\varepsilon_t,v_t)\right|^2 \middle| A_{t-1}\right] \\ &= \sum_{t=1}^{n} E\left[E\left[\left|h_n'\,\delta_n H_{t-1}\psi(\varepsilon_t,v_t)\right|^2 \middle| A_{t-1}\right] \middle| \beta_{n,t-1}\right] + o_p(1) \;\; in\, P_{\theta,n} \\ &= O_p(1) + o_p(1) = O_p(1) \;\; in\, P_{\theta,n}. \end{aligned}$$

The left-hand side of (73) can be rewritten

$$\sum_{t=1}^{n} \left|h_n'\,\delta_n H_{t-1}\right|^2 E\left[E\left[\left|\psi(\varepsilon_t,v_t)\right|^2 \middle| A_{t-1}\right] \middle| \beta_{n,t-1}\right]$$

since H_{t-1} is independent of $\psi(\varepsilon_t,v_t)$ and incorporates only information known at t-1.

Since

$$\begin{aligned} &E\left[E\left[\left|\psi(\varepsilon_t,v_t)\right|^2 \middle| A_{t-1}\right] \middle| \beta_{n,t-1}\right] \\ &= E\left[\left|\psi(\varepsilon_t,v_t)\right|^2 \middle| \beta_{n,t-1}\right] \\ &= \lambda^2 < \infty, \end{aligned}$$

verifying (73) amounts to verifying

$$(75) \qquad \sum_{t=1}^{n} \left| h_n' \delta_n H_{t-1} \right|^2 = O_p(1) \ in\ P_{\theta,n}.$$

We explicitly calculate the weak convergence properties of the left-hand side of (75). The result of this calculation will be important not only to the verification of (75) but also because it will give us the matrix in the quadratic term of the expression for the asymptotic distribution of the log-likelihood ratio, which we can use to distinguish between the LAN and LAMN components of the model.

The left-hand side of (75) is equal to:

$$\sum_{t=1}^{n} h_n' \delta_n H_{t-1} H_{t-1}' \delta_n h_n$$

$$(76) \qquad = \sum_{t=1}^{n} h_n' \begin{bmatrix} \frac{1}{\sqrt{n}} I & 0 \\ 0 & \frac{1}{n} I \end{bmatrix} \begin{bmatrix} Z_{t-1} \\ \Gamma_{t-1} \end{bmatrix} \begin{bmatrix} Z_{t-1}' & \Gamma_{t-1}' \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{n}} I & 0 \\ 0 & \frac{1}{n} I \end{bmatrix} h_n$$

$$= h_n' \sum_{t=1}^{n} \begin{bmatrix} \frac{1}{n} Z_{t-1} Z_{t-1}' & \frac{1}{n^{3/2}} Z_{t-1} \Gamma_{t-1}' \\ \frac{1}{n^{3/2}} \Gamma_{t-1} Z_{t-1}' & \frac{1}{n^2} \Gamma_{t-1} \Gamma_{t-1}' \end{bmatrix} h_n.$$

We shall consider the four blocks of the matrix in the final line of (76) separately, but first we introduce the following notation:

$$\gamma^t(L) \equiv \sum_{j=0}^{t-1} \gamma_j(B, \eta_n) L^j;$$

$$\gamma_r^t(L) \equiv \sum_{j=0}^{t-1-r} \gamma_j(B, \eta_n) \gamma_{j+r}(B, \eta_n) L^j;$$

$$\gamma(L) \equiv \sum_{j=0}^{\infty} \gamma_j(\theta) L^j;$$

$$\gamma_r(L) \equiv \sum_{j=0}^{\infty} \gamma_j(\theta) \gamma_{j+r}(\theta) L^j.$$

First, from Lemma 21 in Jeganathan (1994),

$$(77) \qquad \frac{1}{n}\sum_{t=1}^{n} Z_{t-1} Z_{t-1}' = \gamma(1)^2 \operatorname{var}(\widetilde{u}_1^*, \widetilde{\varepsilon}_1) + o_p(1) \;\; in\, P_{\theta,n},$$

where

$$\widetilde{\varepsilon}_1 = (\varepsilon_1, \ldots, \varepsilon_{2-q}),$$
$$\widetilde{u}_1^* = (u_1^*, \ldots, u_{2-p}^*),$$

and with $(u_t^*, t = 0, \pm 1, \ldots)$ defined as follows (see Lemma 20 of Jeganathan (1994)):

$$u_t^* = \sum_{j=1}^{p} a_j u_{t-j}^* + \sum_{j=1}^{q} b_j \varepsilon_{t-j}^* + \varepsilon_t^*$$

for all $t = 0, \pm 1, \ldots,$ with $(\varepsilon_t^*, t = 0, \pm 1, \ldots)$ forming an iid sequence with $\varepsilon_t^* = \varepsilon_t \;\; \forall t \geq 1.$

Denote the probability limit of the right-hand side of (77) by $J_{ZZ}(\theta)$.

It will now be shown that:

$$(78) \quad \frac{1}{n^2}\sum_{t=1}^{n}\Gamma_{t-1}\Gamma_{t-1}' \Rightarrow \gamma(1)^2(1-\sum_{k=1}^{p}a_k)^2\int_0^1 M_2 M_2',$$

where M_2 is a Brownian motion with covariance matrix $E[v_1 v_1']$, and that

$$(79) \quad \frac{1}{n^{3/2}}\sum_{t=1}^{n} Z_{t-1}\Gamma_{t-1}' = o_p(1) \ \ in\ P_{\theta,n}.$$

We denote the right-hand side of (78) as $J_{\Gamma\Gamma}(\theta)$. To verify (78), note that

$$\begin{aligned}
&\frac{1}{n^2}\sum_{t=1}^{n}\Gamma_{t-1}\Gamma_{t-1}' \\
&= \frac{1}{n^2}\sum_{t=1}^{n}\left[\gamma^t(L)\left[X_t - \sum_{k=1}^{p}a_k^n X_{t-k}\right] - v_t\right] \\
&\qquad \cdot\left[\gamma^t(L)\left[X_t' - \sum_{k=1}^{p}a_k^n X_{t-k}'\right] - v_t'\right].
\end{aligned}$$

Also note that

$$X_t - \sum_{k=1}^{p}a_k^n X_{t-k} = \left(1-\sum_{k=1}^{p}a_k^n\right)X_{t-p} + \sum_{s=0}^{p-1}\left(1-\sum_{k=1}^{s}a_k^n\right)v_{t-s}.$$

Denoting $c_s^n = 1-\sum_{k=1}^{s}a_k^n$, we have

$$\frac{1}{n^2}\sum_{t=1}^{n}\Gamma_{t-1}\Gamma_{t-1}'$$

$$(80)\quad =\frac{1}{n^2}\sum_{t=1}^{n}\left[\gamma^t(L)\left(c_p^n X_{t-p}+\sum_{s=0}^{p-1}c_s^n v_{t-s}\right)-v_t\right] \cdot\left[\gamma^t(L)\left(c_p^n X_{t-p}'+\sum_{s=0}^{p-1}c_s^n v_{t-s}'\right)-v_t'\right].$$

We consider the following six non-redundant components of the right-hand side of (80) separately:

$$(81)\quad \frac{1}{n^2}\sum_{t=1}^{n}\left[\gamma^t(L)c_p^n X_{t-p}\right]\left[\gamma^t(L)c_p^n X_{t-p}'\right]$$

$$(82)\quad \frac{1}{n^2}\sum_{t=1}^{n}\left[\gamma^t(L)c_p^n X_{t-p}\right]\left[\gamma^t(L)\sum_{s=0}^{p-1}c_s^n v_{t-s}'\right]$$

$$(83)\quad \frac{1}{n^2}\sum_{t=1}^{n}\left[\gamma^t(L)c_p^n X_{t-p}\right]v_t'$$

$$(84)\quad \frac{1}{n^2}\sum_{t=1}^{n}\left[\gamma^t(L)\sum_{s=0}^{p-1}c_s^n v_{t-s}\right]\left[\gamma^t(L)\sum_{s=0}^{p-1}c_s^n v_{t-s}'\right]$$

$$(85)\quad \frac{1}{n^2}\sum_{t=1}^{n}\left[\gamma^t(L)\sum_{s=0}^{p-1}c_s^n v_{t-s}\right]v_t'$$

$$(86)\quad \frac{1}{n^2}\sum_{t=1}^{n}v_t v_t'.$$

We begin by showing that (82)-(86) are each $o_p(1)$ in $P_{\theta,n}$.

We begin with (82), which can be rewritten as (using results analogous to those on pp. 978 of Phillips and Solo (1992)):

$$\frac{1}{n}\sum_{s=0}^{p-1} c_s^n c_p^n \sum_{t=1}^{n}\left(\gamma_0^t(1) X_{t-p} v_{t-s}' + \sum_{r=1}^{t}\gamma_r^t(1)\left(X_{t-p}v_{t-s-r}' + X_{t-p-r}v_{t-s}'\right)\right)$$
$$\Rightarrow \sum_{s=0}^{p-1} c_s^n c_p^n \gamma(1)^2 \int_0^1 M_2 dM_2' .$$

because

$$\frac{1}{n}\sum_{s=0}^{p-1} c_s^n c_p^n \sum_{t=1}^{n}\left(\gamma_0^t(1) X_{t-p} v_{t-s}' + \sum_{r=1}^{t}\gamma_r^t(1)\left(X_{t-p}v_{t-s-r}' + X_{t-p-r}v_{t-s}'\right)\right)$$
$$\Rightarrow \sum_{s=0}^{p-1} c_s^n c_p^n \gamma(1)^2 \int_0^1 M_2 dM_2'.$$

We now consider (83). We can rewrite it as:

$$\frac{1}{n^2}\sum_{t=1}^{n}\left(\gamma^t(L) c_p^n X_{t-p}\right) v_t' = o_p(1) \ \ in \, P_{\theta,n}.$$

because

$$\frac{1}{n}\sum_{t=1}^{n}\left(\gamma^t(L) c_p^n X_{t-p}\right) v_t' \Rightarrow \gamma(1) c_p \int_0^1 M_2 dM_2'.$$

We can rewrite (84) as

$$\frac{1}{n}\left[\frac{1}{n}\sum_{s=0}^{p-1}\sum_{r=0}^{p-1} c_s^n c_r^n \sum_{t=1}^{n}\left(\gamma^t(L) v_{t-s}\right)\left(\gamma^t(L) v_{t-r}'\right)\right] = o_p(1) \ \ in \, P_{\theta,n},$$

because, using results on p. 980 of Phillips and Solo (1992), we have

$$\frac{1}{n}\sum_{t=1}^{n}\left(\gamma^{t}(L)v_{t-s}\right)\left(\gamma^{t}(L)v_{t-r}'\right)=\gamma_{r-s}(1)E\left[v_1v_1'\right]\ a.s.\ in\ P_{\theta,n}.$$

We can write (85) as follows:

$$\frac{1}{n}\left[\frac{1}{n}\sum_{t=1}^{n}\left[\sum_{j=0}^{t-1}\sum_{s=0}^{p-1}\gamma_j(B,\eta_n)c_s^n v_{t-j-s}v_t'\right]\right]=o_p(1)\ \ in\,P_{\theta,n},$$

because

$$\frac{1}{n}\sum_{t=1}^{n}\left[\sum_{j=0}^{t-1}\sum_{s=0}^{p-1}\gamma_j(B,\eta_n)c_s^n v_{t-j-s}v_t'\right]=\gamma_0 c_0 E\left[v_1v_1'\right]+o_p(1)\ \ in\,P_{\theta,n}.$$

It is obvious that (86) is $o_p(1)$ in $P_{\theta,n}$. Therefore, in order to verify (78), we need to show that (81) converges in probability measure to $J_{\Gamma\Gamma}(\theta)$. We can write (81) as follows (see Phillips and Solo (1992, p. 978)):

$$\frac{c_p^{n2}}{n^2}\left[\gamma_0^t(1)X_{t-p}X_{t-p}'+\sum_{r=1}^{t-1}\gamma_r^t(1)\left(X_{t-p}X_{t-p-r}'+X_{t-p-r}X_{t-p}'\right)\right]+o_p(1)\ \ in\ P_{\theta,n}$$

$$\Rightarrow c_p^2\gamma(1)^2\int_0^1 M_2M_2'\ \ in\ P_{\theta,n}.$$

We have shown (78) and must now show (79), which can be written

$$
\begin{aligned}
&\frac{1}{n^{3/2}}\sum_{t=1}^{n} Z_{t-1}(\Gamma^{*}_{t-1}-v_t)' \\
&= \frac{1}{n^{3/2}}\sum_{t=1}^{n}\left[\left(\gamma^{t}(L)\left(u_{t-1},\ldots,u_{t-p},\varepsilon_{t-1},\ldots,\varepsilon_{t-q}\right)'\right)\left(\gamma^{t}(L)c_p^{n}X_{t-p}{}'\right)\right] \\
&-\frac{1}{n^{3/2}}\sum_{t=1}^{n}\left[\left(\gamma^{t}(L)\left(u_{t-1},\ldots,u_{t-p},\varepsilon_{t-1},\ldots,\varepsilon_{t-q}\right)'\right)v_t{}'\right] \\
&= o_p(1)+o_p(1) \;\; in\, P_{\theta,n},
\end{aligned}
$$

the second term in this sum being $o_p(1)$ because

$$
\frac{1}{n}\sum_{t=1}^{n}\left[\left(\gamma^{t}(L)\left(u_{t-1},\ldots,u_{t-p},\varepsilon_{t-1},\ldots,\varepsilon_{t-q}\right)'\right)v_t{}'\right] = o_p(1) \;\; in\, P_{\theta,n},
$$

and the first term being $o_p(1)$ because

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\left(\gamma^{t}(L)\left(u_{t-1},\ldots,u_{t-p},\varepsilon_{t-1},\ldots,\varepsilon_{t-q}\right)'\right)\left(\gamma^{t}(L)c_p^{n}X_{t-p}{}'\right) \\
&\Rightarrow \gamma(1)^2 c_p\left(\int_0^1 dM_3M_2{}'+\tau\right)
\end{aligned}
$$

where M_3 is a Brownian motion and τ is a constant.

We then have:

$$
(87) \quad \sum_{t=1}^{n}\begin{bmatrix} \frac{1}{n}Z_{t-1}Z_{t-1}{}' & \frac{1}{n^{3/2}}Z_{t-1}\Gamma_{t-1}{}' \\ \frac{1}{n^{3/2}}\Gamma_{t-1}Z_{t-1}{}' & \frac{1}{n^{2}}\Gamma_{t-1}\Gamma_{t-1}{}' \end{bmatrix} \Rightarrow J(\theta) \equiv \begin{bmatrix} J_{ZZ}(\theta) & 0 \\ 0 & J_{\Gamma\Gamma}(\theta) \end{bmatrix}.
$$

Equation (87) confirms (75) and therefore (73). In order to complete the verification of (68), we must now verify equation (74). Its left-hand side can be written as

$$(88) \quad \sum_{t=1}^{n} |h_n' \delta_n H_{t-1}|^2 E\left[\begin{array}{l} \left(E\left[|\psi(\varepsilon_t, v_t)|^2 | A_{t-1} \right] \right) \\ \cdot I\left(|h_n' \delta_n H_{t-1}| \sqrt{E\left[|\psi(\varepsilon_t, v_t)|^2 | A_{t-1} \right]} > \omega \right) | \beta_{n,t-1} \end{array} \right] \forall \omega > 0.$$

Recall from above that

$$E\left[|\psi(\varepsilon_t, v_t)|^2 | A_{t-1} \right] = \int |\psi(\varepsilon, v_t)|^2 \frac{p(\varepsilon, v_t)}{e(v_t)} d\varepsilon \equiv W^2(v_t),$$

since v_t belongs to the information set A_{t-1} (although not to $\beta_{n,t-1}$).

We can therefore rewrite (88) as

$$(89) \quad \sum_{t=1}^{n} |h_n' \delta_n H_{t-1}|^2 E\left[W^2(v_t) I\left(|h_n' \delta_n H_{t-1}| \cdot |W(v_t)| > \omega \right) | \beta_{n,t-1} \right] \ \forall \omega > 0.$$

We can rewrite (89) as

$$\sum_{t=1}^{n} |h_n' \delta_n H_{t-1}|^2 \int_{R^m} I\left(|W(v)| > \frac{\omega}{|h_n' \delta_n H_{t-1}|} \right) W^2(v) e(v) dv,$$

which is less than or equal to

$$\left(\sum_{t=1}^{n} |h_n' \delta_n H_{t-1}|^2 \right) \int_{R^m} I\left(|W(v)| > \frac{\omega}{\max_{t \in \{1,\dots,n\}} |h_n' \delta_n H_{t-1}|} \right) W^2(v) e(v) dv$$
$$= O_p(1) o_p(1) = o_p(1) \quad in\ P_{\theta,n}.$$

The $O_p(1)$ result is just (75). The $o_p(1)$ result holds because

$$\max_{t\in\{1,\dots,n\}}|h_n'\,\delta_n H_{t-1}| = o_p(1) \ \ in\ P_{\theta,n} \text{ and } \int W^2(v)e(v)dv = \lambda^2 < \infty.$$

To show that $\max_{t\in\{1,\dots,n\}}|h_n'\,\delta_n H_{t-1}| = o_p(1)\ \ in\ P_{\theta,n}$, note that

$$|h_n'\,\delta_n H_{t-1}|^2 = h_n'\begin{bmatrix}\frac{Z_{t-1}Z_{t-1}'}{n} & \frac{Z_{t-1}\Gamma_{t-1}'}{n^{3/2}}\\ \frac{\Gamma_{t-1}Z_{t-1}'}{n^{3/2}} & \frac{\Gamma_{t-1}\Gamma_{t-1}'}{n^2}\end{bmatrix}h_n.$$

Earlier arguments can be used to show that $\max_{t\in\{1,\dots,n\}}|h_n'\,\delta_n H_{t-1}|^2 = o_p(1)\ \ in\ P_{\theta,n}$, so that

$\max_{t\in\{1,\dots,n\}}|h_n'\,\delta_n H_{t-1}| = o_p(1)\ \ in\ P_{\theta,n}$.

So (74) and therefore (71) is verified, and so is (68). We must now verify (69) and (70) to complete the verification of Condition A.3.

We now verify (69). Using the fact that $H_{t-1}^* = H_{t-1} + \begin{bmatrix}0\\ v_t\end{bmatrix}$, verifying (69) involves verifying that the following two conditions hold:

$$(90)\qquad \sum_{t=1}^{n}\int_0^1\!\!\int \left|h_n'\,\delta_n H_{t-1}\left[\psi^*\left(\varepsilon-\kappa h_n'\,\delta_n H_{t-1}^*, v_t\right)-\psi^*\left(\varepsilon, v_t\right)\right]\right|^2 d\varepsilon d\kappa = o_p(1)\ \ in\ P_{\theta,n},$$

and

$$(91)\qquad \sum_{t=1}^{n}\int_0^1\!\!\int \left|h_n'\,\delta_n\begin{bmatrix}0\\ v_t\end{bmatrix}\left[\psi^*\left(\varepsilon-\kappa h_n'\,\delta_n H_{t-1}^*, v_t\right)-\psi^*\left(\varepsilon, v_t\right)\right]\right|^2 d\varepsilon d\kappa = o_p(1)\ \ in\ P_{\theta,n}.$$

To verify (90), we again apply Lemma A.8, defining ξ_{nt}^2 as follows:

$$\xi_{nt}^2 = \int_0^1 \int \left| h_n' \delta_n H_{t-1} \left[\psi^*(\varepsilon - \kappa h_n' \delta_n H_{t-1}^*, v_t) - \psi^*(\varepsilon, v_t) \right] \right|^2 d\varepsilon d\kappa.$$

So we have

$$(92) \sum_{t=1}^{n} E\left[\xi_{nt}^2 \middle| \beta_{n,t-1}\right] = \sum_{t=1}^{n} \left| h_n' \delta_n H_{t-1} \right|^2 \int_0^1 \iint \left[\psi_*(\varepsilon - \kappa h_n' \delta_n H_{t-1}^*, v) - \psi_*(\varepsilon, v) \right]^2 d\varepsilon dv d\kappa$$

where $\psi_*(\varepsilon, v) = \left[\dfrac{\partial p(\varepsilon, v)}{\partial \varepsilon} \right] p(\varepsilon, v)^{-1/2}$.

The right-hand side of (92) is less than or equal to

$$\left(\sum_{t=1}^{n} \left| h_n' \delta_n H_{t-1} \right|^2 \right) \int_0^1 \left(\max_{t \in \{1,\dots,n\}} \iint \left[\psi_*(\varepsilon - \kappa h_n' \delta_n H_{t-1}^*, v) - \psi_*(\varepsilon, v) \right]^2 d\varepsilon dv \right) d\kappa$$

$$= O_p(1) \int_0^1 o_p(1) d\kappa = o_p(1) \ in\, P_{\theta,n}.$$

That the component under the integral is $o_p(1)$ follows from Lemma A.7, the fact that $\int \left| \psi_*(\varepsilon, v) \right|^2 d\varepsilon dv = \lambda^2 < \infty$, and the fact that $\max_{t \in \{1,\dots,n\}} \left| h_n' \delta_n H_{t-1}^* \right| = o_p(1)$ in $P_{\theta,n}$.

This result satisfies both conditions of Lemma A.8, so that (90) follows immediately. We can rewrite (91) as

$$\sum_{t=1}^{n}\left|h_n'\delta_n\begin{bmatrix}0\\ v_t\end{bmatrix}\right|^2\int_0^1\int\left[\psi^*\left(\varepsilon-\kappa h_n'\delta_n H_{t-1}^*,v_t\right)-\psi^*\left(\varepsilon,v_t\right)\right]^2 d\varepsilon d\kappa$$

$$=\sum_{t=1}^{n}\frac{h_{Bn}'v_tv_t'h_{Bn}}{n^2}\int_0^1\iint\left[\psi_*\left(\varepsilon-\kappa h_n'\delta_n H_{t-1}^*,v\right)-\psi_*(\varepsilon,v)\right]^2 d\varepsilon dv d\kappa$$

$$\leq\left(\sum_{t=1}^{n}\frac{h_{Bn}'v_tv_t'h_{Bn}}{n^2}\right)\int_0^1\left(\max_{t\in\{1,\ldots,n\}}\iint\left[\psi_*\left(\varepsilon-\kappa h_n'\delta_n H_{t-1}^*,v\right)-\psi_*(\varepsilon,v)\right]^2 d\varepsilon dv\right)d\kappa$$

$$=o_p(1)\int_0^1 o_p(1)d\kappa=o_p(1)\ \ in\,P_{\theta,n}.$$

So (90), (91), and therefore (69) are verified.

To complete the verification of Condition A.3, it remains to verify (70). This entails showing that

(93) $$\sum_{t=1}^{n}E\left[\left|h_n'\delta_n H_{t-1}\psi\left(\varepsilon_t,v_t\right)\right|^2 I\left(\left|h_n'\delta_n H_{t-1}^*\psi\left(\varepsilon_t,v_t\right)\right|>\omega\right)\middle|A_{t-1}\right]$$
$$=o_p(1)\ \ in\,P_{\theta,n}\ \ \forall\omega>0,$$

and

(94) $$\sum_{t=1}^{n}E\left[\left|h_n'\delta_n\begin{bmatrix}0\\ v_t\end{bmatrix}\psi\left(\varepsilon_t,v_t\right)\right|^2 I\left(\left|h_n'\delta_n H_{t-1}^*\psi\left(\varepsilon_t,v_t\right)\right|>\omega\right)\middle|A_{t-1}\right]$$
$$=o_p(1)\ \ in\,P_{\theta,n}\ \ \forall\omega>0.$$

We begin by showing (93). Again applying Lemma A.8, we set

$$\xi_{nt}^2=E\left[\left|h_n'\delta_n H_{t-1}\psi(\varepsilon_t,v_t)\right|^2 I\left(\left|h_n'\delta_n H_{t-1}^*\psi(\varepsilon_t,v_t)\right|>\omega\right)\middle|A_{t-1}\right],$$

so we have

$$
(95) \quad \begin{aligned} &\sum_{t=1}^{n} E\left[\xi_{nt}^{2} \middle| \beta_{n,t-1}\right] \\ &= \sum_{t=1}^{n} E\left[E\left[\left|h_n'\,\delta_n H_{t-1}\psi(\varepsilon_t, v_t)\right|^2 I\left(\left|h_n'\,\delta_n H_{t-1}^{*}\psi(\varepsilon_t, v_t)\right| > \omega\right) \middle| A_{t-1}\right] \middle| \beta_{n,t-1}\right]. \end{aligned}
$$

Showing that the right-hand side of (95) is $E\left[\left|\psi(\varepsilon_t, v_t)\right|^2 \middle| \beta_{n,t-1}\right] = \lambda^2 < \infty.$ is sufficient to verify (93). It can be rewritten as

$$
\sum_{t=1}^{n} \left|h_n'\,\delta_n H_{t-1}\right|^2 E\left[E\left[\left|\psi(\varepsilon_t, v_t)\right|^2 \cdot I\left(\left|\psi(\varepsilon_t, v_t)\right| > \frac{\omega}{\left|h_n'\,\delta_n H_{t-1}^{*}\right|}\right) \middle| A_{t-1}\right] \middle| \beta_{n,t-1}\right]
$$

which is less than or equal to

$$
\left(\sum_{t=1}^{n} \left|h_n'\,\delta_n H_{t-1}\right|^2\right) E\left[\left|\psi(\varepsilon_t, v_t)\right|^2 \cdot I\left(\left|\psi(\varepsilon_t, v_t)\right| > \frac{\omega}{\max\limits_{t\in[1,n]} \left|h_n'\,\delta_n H_{t-1}^{*}\right|}\right) \middle| \beta_{n,t-1}\right].
$$

The first component is $O_p(1)$ in $P_{\theta,n}$, so we need only show that the second component is $o_p(1)$ in $P_{\theta,n}$ $\forall \omega > 0$. This will follow from the facts that $\max\limits_{t\in\{1,\dots,n\}} \left|h_n'\,\delta_n H_{t-1}^{*}\right| = o_p(1)$ in $P_{\theta,n}$ and that $E\left[\left|\psi(\varepsilon_t, v_t)\right|^2 \middle| \beta_{n,t-1}\right] = \lambda^2 < \infty.$

All that remains in order to verify (70), and to therefore verify Condition A.3, is to verify (94), which can be written as follows:

$$\sum_{t=1}^{n}\left|h_n'\delta_n\begin{bmatrix}0\\ v_t\end{bmatrix}\right|^2 E\left[\left|\psi(\varepsilon_t,v_t)\right|^2 I\left(\left|\psi(\varepsilon_t,v_t)\right|>\frac{\omega}{\left|h_n'\delta_n H_{t-1}^*\right|}\right)\middle|A_{t-1}\right]$$
$$\leq\left(\sum_{t=1}^{n}\left|h_n'\delta_n\begin{bmatrix}0\\ v_t\end{bmatrix}\right|^2\right)E\left[\left|\psi(\varepsilon_t,v_t)\right|^2 I\left(\left|\psi(\varepsilon_t,v_t)\right|>\frac{\omega}{\max_{t\in\{1,\ldots,n\}}\left|h_n'\delta_n H_{t-1}^*\right|}\right)\middle|A_{t-1}\right]$$
$$=o_p(1)\cdot o_p(1)=o_p(1)inP_{\theta,n}.$$

Condition A.4 is satisfied due to equation (108) of Jeganathan (1994), where it is shown that $\int\psi(\varepsilon,v)(p(\varepsilon,v)/e(v))d\varepsilon=0$. As for Condition A.5, we must verify that $f_0(\underline{Y}_0;\theta_n)-f_0(\underline{Y}_0;\theta)=o_p(1)$ in $P_{\theta,n}$ as $\theta_n\to\theta$. The result follows from the following facts:

$$(a)\ f_0(\underline{Y}_0;\theta)=f_0\left(X_1,\ldots,X_n\middle|\varepsilon_{1-q},\ldots,\varepsilon_0;Y_{1-p},\ldots,Y_0;X_{1-p},\ldots,X_0;\theta\right)$$
$$\cdot f_0\left(\varepsilon_{1-q},\ldots,\varepsilon_0;Y_{1-p},\ldots,Y_0;X_{1-p},\ldots,X_0;\theta\right);$$

$$(b)\ f_0\left(X_1,\ldots,X_n\middle|\varepsilon_{1-q},\ldots,\varepsilon_0;Y_{1-p},\ldots,Y_0;X_{1-p},\ldots,X_0;\theta_n\right)$$
$$=f_0\left(X_1,\ldots,X_n\middle|\varepsilon_{1-q},\ldots,\varepsilon_0;Y_{1-p},\ldots,Y_0;X_{1-p},\ldots,X_0;\theta\right)$$

since $\{X_t\}_{t=1}^{n}$ are drawn independently of θ; and

$$(c)\ f_0\left(\varepsilon_{1-q},\ldots,\varepsilon_0;Y_{1-p},\ldots,Y_0;X_{1-p},\ldots,X_0;\theta_n\right)$$
$$-f_0\left(\varepsilon_{1-q},\ldots,\varepsilon_0;Y_{1-p},\ldots,Y_0;X_{1-p},\ldots,X_0;\theta\right)$$
$$=o_p(1)\ in\,P_{\theta,n}$$

as $\theta_n \to \theta$, by an earlier assumption.

This completes the proof of Lemma A.11.

•

We have shown Lemmas A.10 and A.11, proving Theorem 3.1.

•

Proof of Theorem 5.5: We begin by verifying the following lemma:

Lemma A.12: Conditions 5.1-5.3 are satisfied for our model.

Proof: Condition 5.1 holds by earlier assumptions.

To verify Condition 5.2, set $U_{nt}(\theta) = \delta_n H_{t-1}(\theta)$. We first verify that $\{U_{nt}\}_{t=1}^{n}$ so defined satisfy (25) and (26).

The left-hand side of (25) can be rewritten as :

$$\sum_{t=1}^{n} \left| h_n' \delta_n H_{t-1}(\theta) \right|^2 .$$

This quantity can be shown to be $O_p(1)$ in $P_{\theta,n}$ in the same manner as was the left-hand side of (75). The left-hand side of (26) can be written as

$$\max_{t \in \{1,\dots,n\}} \left| h_n' \delta_n H_{t-1}(\theta) \right|,$$

which is $o_p(1)$ in $P_{\theta,n}$. (Recall that in the process of verifying (74), we showed that $\max_{t\in\{1,\ldots,n\}}|h_n'\delta_n H_{t-1}| = o_p(1)$ *in* $P_{\theta,n}$.)

To verify that (24) holds, recall that

$$g_{t-1}(\theta_n) - g_{t-1}(\theta) = h_n'\delta_n H_{t-1}(\eta_n, B)$$

so we must verify that

$$\sum_{t=1}^{n}\left[h_n'\delta_n\left(H_{t-1}(\eta_n,B) - H_{t-1}(\theta)\right)\right]^2 = o_p(1) \ \ in\ P_{\theta,n}$$

or that

$$\begin{aligned}
&h_n'\delta_n\left[\sum_{t=1}^{n}H_{t-1}(\eta_n,B)H_{t-1}(\eta_n,B)'\right]\delta_n h_n + h_n'\delta_n\left[\sum_{t=1}^{n}H_{t-1}(\theta)H_{t-1}(\theta)'\right]\delta_n h_n \\
&-h_n'\delta_n\left[\sum_{t=1}^{n}H_{t-1}(\theta)H_{t-1}(\eta_n,B)'\right]\delta_n h_n - h_n'\delta_n\left[\sum_{t=1}^{n}H_{t-1}(\eta_n,B)H_{t-1}(\theta)'\right]\delta_n h_n \\
&= o_p(1)\ \ in\, P_{\theta,n}.
\end{aligned} \tag{96}$$

Equation (96) will be satisfied, since

$$\delta_n\left[\sum_{t=1}^{n}H_{t-1}(\eta_n,B)H_{t-1}(\eta_n,B)'\right]\delta_n \Rightarrow J(\theta),$$

$$\delta_n\left[\sum_{t=1}^{n}H_{t-1}(\theta)H_{t-1}(\theta)'\right]\delta_n \Rightarrow J(\theta),$$

$$\delta_n\left[\sum_{t=1}^{n}H_{t-1}(\theta)H_{t-1}(\eta_n,B)'\right]\delta_n \Rightarrow J(\theta),\ and$$

$$\delta_n\left[\sum_{t=1}^{n}H_{t-1}(\eta_n,B)H_{t-1}(\theta)'\right]\delta_n \Rightarrow J(\theta).$$

To verify Condition 5.3, set $p+q+m=s$ and $V_{nt}(\theta) = U_{nt}(\theta)$, so that (27) is

$$\sum_{t=1}^{n} \left| u' \delta_n H_{t-1}(\theta_n) - u' \delta_n H_{t-1}(\theta) \right|^2 = o_p(1) \ \ in\ P_{\theta,n}, or$$

$$\begin{aligned}
& u' \delta_n \left[\sum_{t=1}^{n} H_{t-1}(\theta_n) H_{t-1}(\theta_n)' \right] \delta_n u + u' \delta_n \left[\sum_{t=1}^{n} H_{t-1}(\theta) H_{t-1}(\theta)' \right] \delta_n u \\
& - u' \delta_n \left[\sum_{t=1}^{n} H_{t-1}(\theta_n) H_{t-1}(\theta)' \right] \delta_n u - u' \delta_n \left[\sum_{t=1}^{n} H_{t-1}(\theta) H_{t-1}(\theta_n)' \right] \delta_n u \\
& = o_p(1) \ \ in\ P_{\theta,n}.
\end{aligned}$$

This can be verified in the same manner as was (24) above.

To verify (28), set $\delta = 0$, so that we must show that

$$(95) \qquad \max_{t\in\{1,\ldots,n\}} n \left| \delta_n H_{t-1}(\theta) \right|^2 = O_p(1) \ \ in\ P_{\theta,n}.$$

The left-hand side of (95) is

$$\begin{aligned}
& \max_{t\in\{1,\ldots,n\}} n \left| \begin{bmatrix} \frac{1}{\sqrt{n}} I & 0 \\ 0 & \frac{1}{n} I \end{bmatrix} \begin{bmatrix} Z_{t-1}(\theta) \\ \Gamma_{t-1}(\theta) \end{bmatrix} \right|^2 = \max_{t\in\{1,\ldots,n\}} n \left| \begin{matrix} \frac{1}{\sqrt{n}} Z_{t-1}(\theta) \\ \frac{1}{n} \Gamma_{t-1}(\theta) \end{matrix} \right|^2 \\
& = \max_{t\in\{1,\ldots,n\}} n \left[\left(\frac{Z_{t-1}(\theta)' Z_{t-1}(\theta)}{n} \right) + \left(\frac{\Gamma_{t-1}(\theta)' \Gamma_{t-1}(\theta)}{n^2} \right) \right] \\
& = \max_{t\in\{1,\ldots,n\}} \left[Z_{t-1}(\theta)' Z_{t-1}(\theta) + \left(\frac{\Gamma_{t-1}(\theta)' \Gamma_{t-1}(\theta)}{n} \right) \right] \\
& \leq \max_{t\in\{1,\ldots,n\}} \left(Z_{t-1}(\theta)' Z_{t-1}(\theta) \right) + \max_{t\in\{1,\ldots,n\}} \left(\frac{\Gamma_{t-1}(\theta)' \Gamma_{t-1}(\theta)}{n} \right) \\
& = O_p(1) + O_p(1) = O_p(1) \ \ in\ P_{\theta,n}.
\end{aligned}$$

We have verified that Conditions 5.1-5.3 are satisfied for our model. Since the remaining conditions of Proposition 5.4 are also satisfied, equations (29), (30), and (31) hold for our model, so we may now proceed to verify (33).

From (18) and (32), we have

$$(96)\qquad \delta_n^{-1}(\hat{\theta}_n - \tilde{\theta}_n) = \hat{S}_n^{-1} W_n^*(\theta_n^{**}) - \left[\hat{I}_n \frac{S_n(\theta_n^{**})}{\lambda^2}\right]^{-1} \hat{\Delta}_n(\theta_n^{**}).$$

By the definition of $\hat{I}_n$, and using (31), we have

$$(97)\qquad \left[\hat{I}_n \frac{S_n(\theta_n^{**})}{\lambda^2}\right]^{-1} = S_n^{-1}(\theta) + o_p(1) \ \ in\ P_{\theta,n}.$$

Using (30), we have

$$(98)\qquad \hat{\Delta}_n(\theta_n^{**}) = \hat{\Delta}_n(\theta) - S_n(\theta)h_n + o_p(1) \ \ in\ P_{\theta,n}.$$

Combining (98) and (29) gives

$$(99)\qquad \hat{\Delta}_n(\theta_n^{**}) = \Delta_n(\theta) - S_n(\theta)h_n + o_p(1) \ \ in\ P_{\theta,n},$$

so that the second term on the right-hand side of (96) becomes, using (97), (99), and the fact that $\Delta_n(\theta) = W_n(\theta)$,

$$(100)\qquad -S_n^{-1}(\theta)\left[W_n(\theta) - S_n(\theta)h_n\right] + o_p(1) \ \ in\ P_{\theta,n}.$$

By definition,

(101) $\hat{S}_n = S_n(\theta) + o_p(1) \;\; in \; P_{\theta,n},$

while (19) gives us

$$(102) \quad \begin{aligned} W_n^*(\theta_n^{**}) &= W_n(\theta) - \hat{S}_n h_n + o_p(1) \;\; in \; P_{\theta,n} \\ &= W_n(\theta) - S_n(\theta) h_n + o_p(1) \;\; in \; P_{\theta,n}, \end{aligned}$$

the second equality holding due to (101).

Combining (101) and (102), we get

(103) $\hat{S}_n^{-1} W_n^*(\theta_n^{**}) = S_n^{-1}(\theta)\left[W_n(\theta) - S_n(\theta) h_n\right] + o_p(1) \;\; in \; P_{\theta,n}.$

Using (96), (100), and (103), we get

$$\delta_n^{-1}(\tilde{\theta}_n - \hat{\theta}_n) = o_p(1) \;\; in \; P_{\theta,n},$$

our desired result.

•

REFERENCES

Bickel, P.J. 1982. On adaptive estimation. *Annals of Statistics* 10:647-671.

Billingsley, P. 1968. *Convergence of Probability Measures*. New York; John Wiley & Sons.

Fabian, V. and Hannan, J. 1982. On estimation and adaptive estimationfor locally asymptotically normal families. *Z. Wahrsch. verw. Gebiete*. 59:459-478.

Fama, E.F. 1963. Mandelbrot and the stable Paretian hypothesis. *Journal of Business* 36:420-429.

Fama, E.F. 1965. The behaviour of stock market prices. *Journal of Business* 38:34-105.

Fang, K.T., Kotz, S., and Ng, K.W. 1990. *Symmetric Multivariate and Related Distributions*. London; Chapman and Hall.

Ghosh, J.K. 1985. Efficiency of estimates-part I. *Sankhya A* 47:310-325.

Hall, P. and Heyde, C.C. 1980. *Martingale Limit Theory and Its Application*. New York; Academic Press.

Hsieh, D.A. and Manski, C.F. 1987. Monte Carlo evidence on adaptive maximum likelihood estimation of a regression. *Annals of Statistics* 15:541-551.

Jeganathan, P. 1988. Some aspects of asymptotic theory with applications to time series models. University of Michigan, Technical Report #166, September.

Jeganathan, P. 1994. Some aspects of asymptotic theory with applications to time series models. Forthcoming, *Econometric Theory*.

Kreiss, J.-P. 1987a. On adaptive estimation in autoregressive models when there are nuisance functions. *Statistics & Decisions* 5:59-76.

Kreiss, J.-P. 1987b. On adaptive estimation in stationary ARMA processes. *Annals of Statistics* 15:112-133.

LeCam, L. 1960. Locally asymptotically normal families of distributions. *University of California Publications in Statistics* 3:37-98.

LeCam, L. and Yang, G.L. 1990. *Asymptoptics in Statistics - Some Basic Concepts*. New York; Springer-Verlag.

Linton, O. 1993. Adaptive estimation in ARCH models. *Econometric Theory* 9:539-569.

Mandelbrot, B. 1963. The variation of certain speculative prices. *Journal of Business* 36:394-419.

Manski, C.F. 1984. Adaptive estimation of non-linear regression models. *Econometric Reviews* 3:145-210 (including commentary).

Marron, J.S. 1987. Automatic smoothing parameter selection: a survey. *Empirical Economics 13:187-208.*

McDonald, J.B. and White, S.B. 1993. A comparison of some robust, adaptive, and partially adaptive estimators of regression models. *Econometric Reviews* 12:103-124.

McGuirk, A.M., Robertson, J., and Spanos, A. 1993. Modelling exchange rate dynamics: non-linear dependence and thick tails. *Econometric Reviews* 12:3 3-63.

Mitchell, A.F.S. 1989. The information matrix, skewness tensor and $\alpha-$connections for the general multivariate elliptic distribution. *Annals of the Institute of Mathematical Statistics* (Tokyo) 41:289-304.

Mittnik, S. and Rachev, S.T. 1993. Modelling asset returns with alternative stable distributions. *Econometric Reviews* 12:261-330.

Phillips, P.C.B. 1991. Optimal inference in cointegrated systems. *Econometrica* 59:283-306.

Phillips, P.C.B. 1993. Forward market unbiasedness: The case of the Australian dollar since 1984. Cowles Foundation Discussion Paper #1055, August.

Phillips, P.C.B. and Hansen, B.E. 1990. Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57:99-125.

Phillips, P.C.B. and Park, J. 1988. Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated processes. *Journal of the American Statistical Association* 83:111-115.

Phillips, P.C.B. and Solo, V. 1992. Asymptotics for linear processes. *Annals of Statistics* 20:971-1001.

Silverman, B. 1986. *Density Estimation for Statistics and Data Analysis*. London; Chapman and Hall.

Steigerwald, D. 1989. Adaptive estimation in time series regression models. University of California, Santa Barbara, January.

Steigerwald, D. 1992. Adaptive estimation in time series regression models. *Journal of Econometrics* 54:251-276.

White, H. 1984. *Asymptotic Theory for Econometricians*. New York; Academic Press.