TESTING WHEN A PARAMETER IS ON THE BOUNDARY OF THE MAINTAINED HYPOTHESIS

BY

DONALD W. K. ANDREWS

COWLES FOUNDATION PAPER NO. 1021

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2001

http://cowles.econ.yale.edu/

Econometrica, Vol. 69, No. 3 (May, 2001), 683–734

TESTING WHEN A PARAMETER IS ON THE BOUNDARY OF THE MAINTAINED HYPOTHESIS

BY DONALD W. K. ANDREWS[1]

This paper considers testing problems where several of the standard regularity conditions fail to hold. We consider the case where (i) parameter vectors in the null hypothesis may lie on the boundary of the maintained hypothesis and (ii) there may be a nuisance parameter that appears under the alternative hypothesis, but not under the null. The paper establishes the asymptotic null and local alternative distributions of quasi-likelihood ratio, rescaled quasi-likelihood ratio, Wald, and score tests in this case. The results apply to tests based on a wide variety of extremum estimators and apply to a wide variety of models.

Examples treated in the paper are: (i) tests of the null hypothesis of no conditional heteroskedasticity in a GARCH(1, 1) regression model and (ii) tests of the null hypothesis that some random coefficients have variances equal to zero in a random coefficients regression model with (possibly) correlated random coefficients.

KEYWORDS: Asymptotic distribution, boundary, GARCH model, inequality restrictions, random coefficients regression.

1. INTRODUCTION

IN STANDARD TESTING PROBLEMS, parameter values in the null hypothesis are interior points of the maintained hypothesis. For example, the null, alternative, and maintained hypotheses might be $H_0 : \beta_1 = \beta_{1*}$, $H_1 : \beta_1 \neq \beta_{1*}$, and $K : \beta_1 \in R^{p_1}$, respectively, where $K = H_0 \cup H_1$ and β_1 is a subvector of the unknown parameter θ. In addition, in standard testing problems, all parameters are identified under the null and alternative hypotheses.

There are many testing problems of interest where one or both of these features are violated. For example, consider a test of the null hypothesis of no conditional heteroskedasticity in the GARCH(1, 1) model of Bollerslev (1986). The GARCH regression model is $Y_t = X_t'\psi + \varepsilon_t$. The equation for the conditional variance, h_t, of the error ε_t is $h_t = \delta(1-\pi) + \beta_1 \varepsilon_{t-1}^2 + \pi h_{t-1}$ and $h_1 = \delta$, where (β_1, π, δ) are the GARCH moving average (MA), autoregressive (AR), and intercept parameters respectively. In this case, the null hypothesis is $H_0 : \beta_1 = 0$. When $\beta_1 = 0$, we have $h_t = h_{t-1} = \delta$ and the GARCH AR parameter π disappears. Hence, π is unidentified under the null hypothesis. Furthermore, the GARCH MA parameter β_1 must be nonnegative to ensure that the variance is nonnegative. Hence, the alternative and maintained hypotheses are

[1] This paper is dedicated to the memory of the late Dan Nelson. He was a fine person and excellent scholar. He is missed by many in the economics community. The author thanks the Department of Economics at the University of British Columbia for its hospitality while this paper was being written. The author thanks two referees and the co-editor for their comments. The author gratefully acknowledges the research support of the National Science Foundation via Grant Numbers SBR-9410675 and SBR-9730277.

$H_1 : \beta_1 > 0$ and $K : \beta_1 \geq 0$ respectively. Thus, the null value of β_1 is on the boundary of the maintained hypothesis K and the GARCH AR parameter π is unidentified under the null hypothesis.

As a second example where these regularity conditions fail, consider a test of the null hypothesis that b random coefficients have variances that equal zero in a model with random coefficients. Let β_1 denote the sum of the b random coefficient variances that are specified to be zero under the null hypothesis. The null, alternative, and maintained hypotheses are $H_0 : \beta_1 = 0$, $H_1 : \beta > 0$, and $K : \beta_1 \geq 0$ respectively. Hence, under H_0, β_1 is on the boundary of the maintained hypothesis. Suppose that some or all of the b random coefficients whose variances are under test are allowed to be correlated with each other under H_1. Let π_1 denote the vector containing their correlation coefficients. Denote the b vector of random coefficient variances that are under test by $\beta_1 \pi_2$. Here, π_2 is a unit b vector in the nonnegative orthant. Let $\pi = (\pi_1', \pi_2')'$. Under the alternative hypothesis, β_1 and π together specify the variances and covariances of the random coefficients under test. Under the null hypothesis, the parameter π is not identified.

In this paper, we provide general asymptotic results that cover testing problems of the above sort. We specify a set of high level conditions under which the asymptotic null distributions of quasi-likelihood ratio (QLR), rescaled quasi-likelihood ratio (RQLR), Wald, and score tests are determined. We provide several sets of more primitive sufficient conditions that imply the high level conditions. We verify the latter conditions in the two examples described above (using a random coefficients regression model in the second example). We show that the asymptotic distributions of the test statistics under local alternatives can be established using the same set of general results that are used under the null. The results given here utilize and extend the estimation results of Andrews (1999) and Andrews (1997), which we refer to as E1 and E2 respectively.

In the GARCH example, the asymptotic null distribution of the test statistics is found to be nuisance parameter free (under appropriate assumptions). Asymptotic critical values are provided. In the random coefficient regression example, the asymptotic null distribution of the test statistics is found to be nuisance parameter dependent, but critical values and p-values can be simulated.

We now give a brief overview of the method used to obtain the asymptotic null distributions of the test statistics. Let the estimator objective function be $\ell_T(\theta, \pi)$, where (θ, π) are parameters that lie in the maintained hypothesis parameter space $\Theta \times \Pi$ and T denotes the sample size. The null hypothesis can be written as $H_0 : \theta \in \Theta_0$, where Θ_0 is a subset of Θ. The objective function could be a quasi-log likelihood, least squares, generalized method of moments, or semiparametric objective function, among others. The results allow for objective functions that are smooth or nonsmooth functions of the parameters (θ, π). The results allow for nonlinear models with nontrending or deterministically trending data, as well as linear models with nontrending, deterministically trending, and/or stochastically trending data. See E1 and E2 for examples that exhibit these different features.

The basic idea is to (i) approximate $\ell_T(\theta, \pi)$ by a quadratic function of θ whose coefficients depend on π; (ii) show that the approximation holds uniformly over $\pi \in \Pi$; (iii) approximate the unrestricted and restricted parameter spaces, Θ and Θ_0, by cones, as in Chernoff (1954) and E1; (iv) determine the asymptotic distributions of the (suitably normalized) unrestricted and restricted estimators of θ given π and of the estimator objective function evaluated at these two estimators *as stochastic processes indexed by* $\pi \in \Pi$; and (v) obtain the asymptotic null distributions of the test statistics by writing them as continuous functions of the (normalized) estimators and/or the objective function evaluated at these estimators and applying the continuous mapping theorem.

The asymptotic distributions of the test statistics are given by the differences between the suprema over $\pi \in \Pi$ of two stochastic processes indexed by π that are each a quadratic form in a random vector that minimizes a stochastic quadratic function over a cone. For the Wald and score statistics and in some cases for the QLR and RQLR statistics, the second quadratic form is degenerate and equals zero. The asymptotic distributions may depend on (estimable) nuisance parameters. Critical values can be obtained straightforwardly by simulation given consistent estimates of any unknown nuisance parameters.

Some examples that are covered by the general results, but are not discussed in this paper, include: tests of the null of parameter stability against the alternative of one-sided structural change; tests of white noise against serial correlation in ARMA(1, 1) models with positive autocorrelation, one-sided tests of the significance of Box-Cox transformed regressors; one-sided tests of threshold effects in threshold models, such as threshold autoregressive models; tests of the null that random coefficient variances are zero in nonlinear models, such as the conditional probit model of Hausman and Wise (1978); tests that random coefficient variances are zero when the random coefficients may be correlated across time (see Rosenberg (1973) and Watson and Engle (1985)); tests for the presence of conditional heteroskedasticity in models other than the GARCH(1, 1) model (e.g., see Bollerslev, Engle, and Nelson (1994)); and tests in the examples in E1 and E2. The latter include tests of a unit root in a Dickey-Fuller regression model with time trend and autoregressive root restricted to be nonnegative and less than or equal to one respectively; tests of a GARCH$(1, q^*)$ model against a GARCH$(1, q^* + p_1)$ model; tests of equality and/or inequality restrictions in a regression model with integrated regressors, median regression model, or partially linear regression model; tests of zero variance of measurement errors and/or random effects, in a multinomial response model estimated by the method of simulated moments of McFadden (1989).

There are numerous antecedents in the literature to the approach taken here. For example, the use of a quadratic approximation to the estimator objective function, rather than the reliance on first order conditions, has been made by Chernoff (1954), LeCam (1960), Jeganathan (1982), Pollard (1985), Pakes and Pollard (1989), Geyer (1994), van der Vaart and Wellner (1996), and others. Our treatment of nonsmooth estimator objective functions is via stochastic equicontinuity or stochastic differentiability conditions, as in Pollard (1985), Pakes and Pollard (1989), Andrews (1994a, b), Geyer (1994), Newey and McFadden (1994), and van der Vaart and Wellner (1996). Several papers in the literature consider tests when parameter vectors in the null are on the boundary of the maintained hypothesis. These include Chernoff (1954), Bartholomew (1959), Perlman (1969), Chant (1974), Shapiro (1985), Self and Liang (1987), Gourieroux and Monfort (1989, Ch. 21), Andrews (1996, 1998), and King and Wu (1997). See Barlow, Bartholomew, Bremner, and Brunk (1972) and Wu and King (1994) for further references. Several papers consider tests when a nuisance parameter appears under the alternative hypothesis, but not under the null. These include Davies (1977, 1987), King and Shively (1993), Andrews and Ploberger (1994, 1995), and Hansen (1996), among others.

For a review of the testing results in the literature for the GARCH example, see Bollerslev, Engle, and Nelson (1994). For the random coefficients example, see Brooks and King (1994). The results in the literature do not cover the QLR, RQLR, Wald, or score tests considered here.

To compare the results of this paper to papers in the literature, we note the main features of the present paper. They are: (i) parameter vectors in the null may be on the boundary of the maintained hypothesis, (ii) there may be a nuisance parameter that appears under the alternative but not under the null, (iii) the boundary of the maintained hypothesis may be curved and/or kinked, (iv) the estimator objective function need not be defined in a full neighborhood of the true parameter (which is necessary to cover random coefficient models), (v) the estimator objective function may be smooth or nonsmooth, (vi) the estimator objective function can be a quasi-log likelihood, least squares, generalized method of moments, minimum distance, or semiparametric objective function, among others, (vii) the data may contain deterministic and/or stochastic trends in linear models, and (viii) rescaled quasi-likelihood ratio tests are analyzed. None of the papers in the literature allow for more than two of these features simultaneously and most allow for just one.

The present paper provides local power results, but does not establish the asymptotic admissibility of the tests considered. It may be possible to do so under suitable assumptions, perhaps using the methods of Andrews and Ploberger (1994) and Andrews (1996). It also may be possible to introduce a class of tests that have some weighted average power optimality properties, along the lines of Andrews and Ploberger (1994) and Andrews (1998). These are topics for future research. We note that, in cases where the restrictions on the parameter space arise from prior information, tests that utilize this information, such as the tests considered here, have a considerable power advantage over tests that do not. For example, see the power comparisons in Andrews (1998, Table 2).

The remainder of this paper is organized as follows. Section 2 introduces the GARCH and random coefficient examples. Section 3 determines the asymptotic behavior of the unrestricted extremum estimator when there is an unidentified-nuisance parameter. Section 4 applies these results to the two examples. Section 5 defines the QLR and RQLR test statistics, determines their asymptotic null distributions, and applies the results to the two examples. Sections 6 and 7 do likewise for the Wald and score tests. Section 8 establishes the asymptotic distributions of the test statistics under local alternatives. An Appendix contains proofs of results given in the paper.

All limits below are taken "as $T \to \infty$." Let $o_{p\pi}(1)$, $O_{p\pi}(1)$, and $o_{\pi}(1)$ denote terms that are $o_p(1)$, $O_p(1)$, and $o(1)$, respectively, uniformly over $\pi \in \Pi$. Thus, $X_{T\pi} = o_{p\pi}(1)$ means that $\sup_{\pi \in \Pi} \|X_{T\pi}\| = o_p(1)$, where $\|\bullet\|$ denotes the Euclidean norm. Let "wp $\to 1$" abbreviate "with probability that goes to one as $T \to \infty$." Let "for all $\gamma_T \to 0$" abbreviate "for all sequences of positive scalar constants $\{\gamma_T : T \geq 1\}$ for which $\gamma_T \to 0$." Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and largest eigenvalues, respectively, of a matrix A. Let $\partial\Lambda$ denote the boundary and $\mathrm{cl}(\Lambda)$ denote the closure of a set Λ. Let $S(\theta, \varepsilon)$ denote an open sphere centered at θ with radius ε. Let $C(\theta, \varepsilon)$ denote an open cube centered at θ with sides of length 2ε. Let $:=$ denote "equals by definition." Let $\simeq$ denote equality in distribution.

Let $\Rightarrow$ denote weak convergence of a sequence of stochastic processes indexed by $\pi \in \Pi$ for some space Π. The definition of weak convergence of R^v-valued functions on Π requires the specification of a metric d on the space $\mathscr{E}_v$ of R^v-valued functions on Π. We take d to be the uniform metric. The literature contains several definitions of weak convergence. We use any of the definitions that is compatible with the use of the uniform metric and for which the continuous mapping theorem holds. These include the definitions employed by Pollard (1984, p. 65), Pollard (1990, p. 44), and van der Vaart and Wellner (1996, p. 17). The continuous mapping theorems that correspond to these definitions are given by Pollard (1984, p. 70), Pollard (1990, p. 46), and van der Vaart and Wellner (1996, Thm. 1.3.6, p. 20).

2. EXAMPLES

2.1. *GARCH Example*

In this example, we consider testing the null hypothesis of no conditional heteroskedasticity in a GARCH(1,1) regression model. The null hypothesis is $H_0 : \beta_1 = 0$, where β_1 is the GARCH-MA coefficient.

The observed random variables are $\{(Y_t, X_t) : 1 \leq t \leq T\}$. The model used to generate a quasi-likelihood function is the normal GARCH(1,1) model:

$$(2.1) \qquad Y_t = X_t'\psi + h_t^*(\theta, \pi)^{1/2} z_t \quad \text{for} \quad t = 1, \ldots, T,$$

$$h_t^*(\theta, \pi) := \delta(1 - \pi) + \beta_1 e_{t-1}^2(\theta) + \pi h_{t-1}^*(\theta, \pi) \quad \text{for} \quad t = 2, \ldots, T,$$

$$e_t(\theta) = Y_t - X_t'\psi, \qquad \theta := (\beta_1, \delta, \psi')',$$

and $\{z_t : t = 1, \dots, T\}$ are iid $N(0,1)$ and are independent of $\{X_t : t = 1, \dots, T\}$, where ψ, $X_t \in R^r$ and $h_t^*(\theta, \pi), z_t, e_t(\theta), \beta_1, \delta, \pi \in R$. The initial condition $h_1^*(\theta, \pi)$ is an arbitrary nonnegative function of (θ, π, Y_1, X_1) that satisfies $\sup_{\theta \in \Theta, \pi \in \Pi} h_1^*(\theta, \pi) < \infty$ a.s., where the parameter spaces Θ and Π are defined below. For the QLR test considered below, however, the initial condition must be such that $h_1^*(\theta_0, \pi) = \delta_0$, where θ_0 and δ_0 are the true parameter values defined in (2.3) below. The choice $h_1^*(\theta, \pi) = \delta$ suffices. The true process generating the data is specified in (2.3) below and does not necessarily satisfy the model. For example, the innovations need not be iid normal.

The parameter space Θ is a compact subset of R^s that restricts the GARCH−MA parameter, β_1, to be nonnegative and bounds the conditional variance intercept parameter, δ, away from zero. The parameter space Π is a compact subset of $[0,1]$ that bounds π away from one.

$$\begin{aligned}(2.2)\qquad & \Theta := \{\theta \in R^s : \theta = (\beta_1, \delta, \psi')', 0 \le \beta_1 \le \beta_{1u}, 0 < \delta_\ell \le \delta \le \delta_u, \text{ and} \\ & \quad \psi_\ell \le \psi \le \psi_u, \text{ where } \beta_{1u}, \delta_\ell, \delta_u, \psi_\ell, \text{ and } \psi_u \text{ are some known} \\ & \quad \text{finite constants or constant vectors}\}. \\ & \Pi \subset \{\pi \in [0, \pi_u] : \pi_u \text{ is a constant less than one}\}.\end{aligned}$$

(The vector inequality involving ψ is an element by element inequality.) Note that the parameter space need not restrict the GARCH parameters to be values that generate a stationary process. Under the null hypothesis, however, the true process is stationary.

We derive asymptotic results for the case where the null hypothesis is true. The true parameter vector under the null hypothesis is $\theta_0 := (\beta_{10}, \delta_0, \psi_0')' = (0, \delta_0, \psi_0')' \in R^s$. The true process generating $\{(Y_t, X_t) : 1 \le t \le T\}$ is

$$\begin{aligned}(2.3)\qquad & Y_t := X_t'\psi_0 + \varepsilon_t, \qquad \varepsilon_t := \delta_0^{1/2} z_t, \\ & \{(z_t, X_t) : t = \dots, 0, 1, \dots\} \quad \text{are stationary and ergodic}, \\ & E(z_t | \mathscr{F}_{t-1}) = 0 \quad \text{a.s.}, \qquad E(z_t^2 | \mathscr{F}_{t-1}) = 1 \quad \text{a.s.} \\ & \qquad \text{where} \quad \mathscr{F}_t = \sigma(z_t, z_{t-1}, \dots, X_{t+1}, X_t, \dots), \\ & P(z_t^2 = 1) \ne 1, \qquad E(z_t^4 | \mathscr{F}_{t-1}) = \kappa < \infty \quad \text{a.s. for some constant } \kappa, \\ & EX_tX_t' \text{ is positive definite, and} \\ & E\left(1 + \|X_t\|^2\right)\left(z_{t-j}^4 + \|X_{t-j}\|^4\right) < \infty \qquad \forall j \ge 1.\end{aligned}$$

For example, the last moment condition holds if $E\|X_t\|^6 < \infty$ and $Ez_t^6 < \infty$. The regressor vector X_t need not be strictly exogenous and the innovation z_t need not have distribution that is normal or symmetric about zero.

We assume that $\theta_0 \in \Theta$ and that each subvector of θ_0 satisfies the inequalities imposed by Θ strictly except β_{10}, which equals 0 and causes θ_0 to be on the boundary of Θ. (It is possible to show that the testing results given below are invariant to the regression parameter ψ_0 being on a boundary, but for brevity we do not do so here.)

In contrast to the GARCH example of E2, the GARCH example considered here is one in which the true process does not exhibit conditional heteroskedasticity. This causes a nuisance parameter π to appear that is not identified, which complicates the analysis. But, it allows us to consider tests for the existence of conditional heteroskedasticity, which are of considerable interest.

2.2. *Random Coefficient Example*

The second example is a random coefficient regression model. This model was first considered by Hildreth and Houck (1968). We are interested in testing the null hypothesis that some, or perhaps all, of the random coefficient variances are zero. We allow for the possibility that the random coefficients are correlated. For example, this is a realistic feature when the random coefficients are taste parameters of an individual that are randomly distributed across some population, because one would expect the tastes of a given individual to be correlated. The random coefficient model considered in E1 is less general than the one considered here, because it does not allow for correlation between the random coefficients.

The model is

$$(2.4)\qquad \begin{aligned} Y_t &= \psi_2 + X_t'\gamma_t + \delta_2^{1/2}\varepsilon_t \\ &= \psi_2 + X_t'\psi_1 + (\delta_2^{1/2}\varepsilon_t + X_t'\Omega^{1/2}(\beta_1, \delta_1, \pi)\eta_t), \qquad \text{where} \\ \gamma_t &:= \psi_1 + \Omega^{1/2}(\beta_1, \delta_1, \pi)\eta_t. \end{aligned}$$

The vector $\gamma_t \in R^{b+c}$ is the random coefficient vector. The observed variables are $\{(Y_t, X_t): \leq t \leq T\}$. The random variables $\{(Y_t, X_t, \varepsilon_t, \eta_t): \leq t \leq T\}$ are iid. The regressors are $X_t := (X_{1t}', X_{2t}')'$, where $X_{1t} \in R^b$ and $X_{2t} \in R^c$. Without loss of generality, X_{1t} consists of the regressors whose random coefficients have zero variance under the null and X_{2t} consists of the remaining regressors. The random variables $\eta_t \in R^{b+c}$ and $\varepsilon_t \in R$ are unobserved errors. We consider the quasi-likelihood function that is generated by $(\varepsilon_t, \eta_t')' \sim N(\mathbf{0}, I_{1+b+c})$, but the true process need not satisfy this condition. We assume that the true process is such that $E\varepsilon_t = 0$, $E\varepsilon_t^2 = 1$, $E(\eta_t \mid X_t) = \mathbf{0}$ a.s., $E(\eta_t\eta_t' \mid X_t) = I_{b+c}$ a.s., and $E(\eta_t\varepsilon_t \mid X_t) = \mathbf{0}$ a.s.

The $(b+c)\times(b+c)$ covariance matrix of the random coefficients is $\Omega(\beta_1, \delta_1, \pi)$. It is of the form

$$(2.5)\qquad \Omega(\beta_1, \delta_1, \pi) := \begin{bmatrix} \beta_1\Omega_1(\pi) & \mathbf{0} \\ \mathbf{0} & \Omega_2(\delta_1) \end{bmatrix}.$$

The parameter $\beta_1 \geq 0$ equals the sum of the random coefficient variances whose values are zero under the null hypothesis. Because we derive asymptotic results for the case when the null hypothesis is true, the true value of β_1, β_{10}, is zero. The $b \times b$ matrix $\beta_1 \Omega_1(\pi)$ is the covariance matrix of the random coefficients that are under test. It depends on the vector $\pi = (\pi_1', \pi_2')' \in R^{d+b}$. The vector $\pi_1 \in R^d$ contains all of the correlation parameters between the random coefficients on X_{1t} that are allowed to be nonzero under H_1. Its dimension d lies between zero and $b(b-1)/2$. The vector π_2 is a unit b vector in the nonnegative orthant, i.e., a direction vector. The b vector of random coefficient variances on X_{1t} is $\beta_1 \pi_2$. We choose this parameterization of the covariance matrix of the random coefficients, because it yields an estimator objective function that is well-behaved—its (generalized) first and second derivatives are continuous in β_1 at $\beta_1 = 0$. Some other parameterizations do not do so.

For $i, j = 1, \ldots, b$, the (i, j) element of $\Omega_1(\pi)$ is

$$[\Omega_1(\pi)]_{ij} := \pi_{2i}^{1/2} \pi_{2j}^{1/2} \rho_{ij}, \qquad \text{where} \tag{2.6}$$

$$\pi_1 := (\pi_{11}, \ldots, \pi_{1d})',$$

$$\pi_2 := (\pi_{21}, \ldots, \pi_{2b})',$$

and $\rho_{ij} := 1$ if $i = j$, $\rho_{ij} := \pi_{1\ell}$ for some $\ell \leq d$ if the correlation between the random coefficients on the ith and jth elements of X_{1t} is allowed to be nonzero under the maintained hypothesis, and $\rho_{ij} := 0$ if the correlation between the random coefficients on the ith and jth elements of X_{1t} is zero under the maintained hypothesis.

The vector $\delta_1 \in R^g$ contains any random coefficient variances on elements of X_{2t} plus any correlation parameters between random coefficients on elements of X_{2t}. The coefficients on X_{2t} need not be random. If none of them are, then $\Omega_2(\delta_1) = \mathbf{0}$ and the parameter δ_1 does not appear. The block diagonality of $\Omega(\beta_1, \delta_1, \pi)$ reflects our assumption that the correlations between random coefficients on X_{1t} and those on X_{2t} are specified to be zero under the maintained hypothesis and, hence, are not estimated. Whether or not this assumption holds, the tests considered below have correct significance level asymptotically (because the correlations are necesssarily zero under the null hypothesis). This assumption affects the form of the test statistics, however, and affects the power of the tests positively or negatively, depending on how small or large the correlations are if they are nonzero. Note that nonblock diagonality of $\Omega(\beta_1, \delta_2, \pi)$ would cause theoretical problems because the off-diagonal parameters would be of the form $\beta_1^{1/2}\kappa$, where κ is a function of δ_1 and π, the partial derivative of $\beta_1^{1/2}\kappa$ with respect to β_1 is $(1/2)\beta_1^{-1/2}\kappa$, and the latter equals infinity at $\beta_1 = 0$.

The parameter δ_2 is the idiosyncratic error variance. The parameter $\psi_1 \in R^{b+c}$ is the deterministic part of the regression coefficients. The parameter $\psi_2 \in R$ is the regression intercept. The vectors $\theta := (\beta_1, \delta_1', \delta_2, \psi_1', \psi_2)'$ and π are the unknown parameters to be estimated. The parameter space Θ of θ is

$$\Theta := \{\theta \in R^s : \theta = (\beta_1, \delta_1', \delta_2, \psi_1', \psi_2)', 0 \le \beta \le \beta_{1u}, \delta_{j\ell} \le \delta_j \le \delta_{ju} \text{ and } \psi_{j\ell} \le \psi_j \le \psi_{ju} \text{ for } j = 1,2\} \tag{2.7}$$

for some known finite constants or vectors $\beta_{1u}, \delta_{j\ell}, \delta_{ju}, \psi_{j\ell}, \psi_{ju}$ for $j = 1,2$, where $\delta_{2\ell} > 0$, the lower bound on each variance parameter in δ_1 is greater than or equal to zero, and the lower and upper bounds on each correlation parameter in δ_1 are between -1 and 1 respectively. The parameter space Π of π is a compact subset of $(-1,1)^d \times \mathscr{U}_b^+$, where $\mathscr{U}_b^+$ denotes the set of unit vectors in R^b that are in the nonnegative orthant.

The true parameter vector θ_0 is

$$\theta_0 := (\beta_{10}, \delta_{10}', \delta_{20}, \psi_{10}', \psi_{20})' = (0, \delta_{10}', \delta_{20}, \psi_{10}', \psi_{20})', \tag{2.8}$$

where none of the restrictions defining Θ are binding at θ_0 except $\beta_{10} = 0$.

The above specification of the model requires that all of the variances that are not under test are positive. That is, each variance parameter in δ_1 is positive. If one is not sure that this is true, then one could carry out a sequence of tests as follows. First, order the variances of all of the random coefficients that might possibly be zero, say $0 \le \omega_1^2 \le \cdots \le \omega_{\bar{b}}^2$ for some $\bar{b} < \infty$. Define $\beta_{1,k} = \sum_{j=1}^k \omega_j^2$ for $k = 1, \ldots, \bar{b}$. Next, for $k = \bar{b}, \ldots, 1$, test sequentially $H_{0,k} : \beta_{1,k} = 0$ versus $H_{1,k} : \beta_{1,k} > 0$, where one continues in the sequence until a test fails to reject the null. At each stage of the testing sequence, all of the coefficient variances that are not under test are positive, as assumed above, provided none of the preceding tests have made an error.

If the random coefficient model of interest specifies the random coefficients under test to be *uncorrelated*, then the parameter π_1 disappears from the model. In this case, an alternative parameterization can be employed, as in Andrews (1999). One can define the vector of variances of the random coefficients on X_{1t} to be $\beta_1 \in (R^+)^b$, rather than $\beta_1 \pi_2$. In this case, the hypotheses of interest are $H_0 : \beta_1 = \mathbf{0}$ and $H_1 : \beta_1 \ne \mathbf{0}$ and $\beta_1 \ge \mathbf{0}$. With this parameterization, no parameter π_2 appears. This parameterization has the feature that there is no parameter π that appears under the alternative but not under the null. It still has the feature that the parameter β_1 lies on the boundary of the maintained hypothesis. This type of parameterization is not appropriate if there is correlation between the random coefficients on X_{1t}, because the (generalized) first and second derivatives of the quasi-log likelihood function are not continuous at $\beta_1 = \mathbf{0}$.

The model considered in (2.4)–(2.8) could be generalized by adding a third diagonal block to the variance matrix $\Omega(\beta_1, \delta_1, \pi)$ that corresponds to parameters whose variances are not under test, but whose values are zero. This block would be of the form $\beta_2 \Omega_2(\pi)$, where $\beta_2 \ge 0$ is the sum of the random coefficient variances in question, the true value of β_2 is zero, $\Omega_2(\pi)$ is of the same form as $\Omega_1(\pi)$, and π is elongated to include additional parameters that appear in $\Omega_2(\pi)$. In this case, β_2 is a parameter that is not under test but is on the boundary of the parameter space. For brevity, we do not discuss this case further.

3. THE UNRESTRICTED EXTREMUM ESTIMATOR

3.1. *Notation and the Hypotheses of Interest*

The data matrix for sample size T is $\mathbf{Y}_T$ for $T = 1, 2, \ldots$. We consider an estimator objective function $\ell_T(\theta, \pi)$ that depends on $\mathbf{Y}_T$ and on the parameters θ and π. The parameter spaces for θ and π are Θ and Π, where $\Theta \subset R^s$ and Π is some space (usually a subset of Euclidean space). Below we consider estimators and tests based on $\ell_T(\theta, \pi)$. Although it is convenient to view $\ell_T(\theta, \pi)$ as a quasi-log likelihood function, the results below do not require this. The function $\ell_T(\theta, \pi)$ could be any objective function desired, such as the negative of a LS, GMM, minimum distance, or semiparametric objective function.

We adopt the same basic notation as in E1 except that we allow the estimator objective function to depend on a nuisance parameter π that is unidentified when the true parameter is in the null hypothesis. Much of the discussion of the assumptions and results given in E1 is applicable in this section too. For this reason, we keep the discussion here as brief as possible.

The null, alternative, and maintained hypotheses that we consider are

$$(3.1) \qquad H_0 : \theta \in \Theta_0, \pi \in \Pi; \quad H_1 : \theta \in \Theta_1, \pi \in \Pi; \quad \text{and} \quad K : \theta \in \Theta, \pi \in \Pi;$$

respectively, where $\Theta_0 \subset \Theta \subset R^s$ and $\Theta_1 := \Theta / \Theta_0$. The null hypothesis is a point null hypothesis of the form $H_0 : \beta_1 = \beta_{1*}$, where β_1 is a subvector of θ (see Assumption 9 in Section 5 below).

We consider the case where the null hypothesis exhibits the property that $\ell_T(\theta, \pi)$ does not depend on π when θ is the null hypothesis. In consequence, π is unidentified under the null hypothesis. Actually, in some time series contexts of interest, $\ell_T(\theta_0, \pi)$ *does* depend on π, due to the effect of initial conditions, but π is still asymptotically unidentified. This has no effect on the Wald and score tests and our results cover this case. It does, however, have an effect on the QLR test and the QLR statistic has the appropriate asymptotic null distribution only if $\ell_T(\theta, \pi)$ does not depend on π for $\theta \in \Theta_0$.

The testing scenario considered here includes the standard case where no parameter π appears that is unidentified under the null. To cover such cases, one takes the parameter space Π of π to include a single point.

Let θ_0 denote the pseudo-true value of the parameter θ. That is, θ_0 is the probability limit of the unrestricted and restricted estimators introduced below. We assume that θ_0 is in the null hypothesis, i.e., $\theta_0 \in \Theta_0$, because we are interested in the asymptotic null distributions of various test statistics. If the model is correctly specified, then θ_0 denotes the true value. Even if the model is misspecified, however, it may be meaningful to speak of the true value of θ or of the true value of some subvector of θ, such as β_1. For example, in both examples considered above, if the innovations or errors are not normal, then the model is misspecified, but the true value of θ is still well defined. Thus, even if the model is misspecified, the null hypothesis may specify a restriction that is of interest.

3.2. *Definition of the Unrestricted Estimator and Consistency*

We now define the unrestricted extremum estimator $\hat{\theta}_\pi$ of θ for given $\pi \in \Pi$. By definition, $\hat{\theta}_\pi \in \Theta \;\; \forall \pi \in \Pi$ and

$$(3.2) \qquad \ell_T\left(\hat{\theta}_\pi, \pi\right) = \sup_{\theta \in \Theta} \ell_T(\theta, \pi) + o_{p\pi}(1) \qquad \forall \pi \in \Pi.$$

The $o_{p\pi}(1)$ term is included in (3.2) (and in various definitions below) to indicate that the supremum does not need to be obtained exactly.

We assume the following.

ASSUMPTION 1: $\hat{\theta}_\pi = \theta_0 + o_{p\pi}(1)$.

Assumption 1 typically holds because $\ell_T(\theta, \pi)$ does not depend on π when θ is in the null hypothesis (at least up to an asymptotically negligible term). A sufficient condition for Assumption 1 for models with nontrending data is the following.

ASSUMPTION 1*: (a) *For some nonrandom function* $\ell(\theta, \pi): \Theta \times \Pi \to R$,

$$\sup_{\theta \in \Theta,\, \pi \in \Pi} \left| T^{-1} \ell_T(\theta, \pi) - \ell(\theta, \pi) \right| \to^p 0.$$

(b) *For all* $\varepsilon > 0$, $\sup_{\theta \in \Theta / S(\theta_0, \varepsilon),\, \pi \in \Pi} \ell(\theta, \pi) < \ell(\theta_0)$, *where* $\ell(\theta_0) := \ell(\theta_0, \pi)$ *does not depend on* π *and* $\Theta / S(\theta_0, \varepsilon)$ *denotes all vectors* θ *in* Θ *but not in* $S(\theta_0, \varepsilon)$.

The sufficiency of Assumption 1* for Assumption 1 follows from Lemma A-1 of Andrews (1993).

Note that here and below a superscript *, 2*, 3*,... on an assumption denotes that the assumption is sufficient (sometimes only in the presence of other assumptions) for the unsuperscripted assumption.

3.3. *Quadratic Approximation of the Objective Function*

The objective function is assumed to have a quadratic expansion in θ about θ_0 for each $\pi \in \Pi$:

$$
\begin{aligned}
(3.3)\qquad \ell_T(\theta,\pi) &:= \ell_T(\theta_0,\pi) + D\ell_T(\theta_0,\pi)'(\theta-\theta_0) \\
&\quad + \frac{1}{2}(\theta-\theta_0)'D^2\ell_T(\theta_0,\pi)(\theta-\theta_0) + R_T(\theta,\pi) \\
&= \ell_T(\theta_0,\pi) + \frac{1}{2}Z'_{T\pi}\mathcal{J}_{T\pi}Z_{T\pi} \\
&\quad - \frac{1}{2}q_T(B_T(\theta-\theta_0),\pi) + R_T(\theta,\pi), \qquad \text{where}
\end{aligned}
$$

$$
\mathcal{J}_{T\pi} := -B_T^{-1\prime}D^2\ell_T(\theta_0,\pi)B_T^{-1}, \quad Z_{T\pi} := \mathcal{J}_{T\pi}^{-1}B_T^{-1\prime}D\ell_T(\theta_0,\pi), \quad \text{and}
$$

$$
q_T(\lambda,\pi) := (\lambda - Z_{T\pi})'\mathcal{J}_{T\pi}(\lambda - Z_{T\pi}) \qquad \text{for} \qquad \lambda \in R^s.
$$

We allow $\ell_T(\theta_0,\pi)$ to depend on π to allow for the possible effect of initial conditions. Note that even when $\ell_T(\theta_0,\pi)$ does not depend on π, the generalized derivatives $D\ell_T(\theta_0,\pi)$ and $D^2\ell_T(\theta_0,\pi)$ of $\ell_T(\theta_0,\pi)$ usually depend on π, because $\ell_T(\theta,\pi)$ usually depends on π for θ not in the null hypothesis but arbitrarily close to θ_0.

The terms in the quadratic expansion are assumed to satisfy:

ASSUMPTION 2: *For all* $0 < \gamma < \infty$, $\sup_{\theta \in \Theta : \|B_T(\theta-\theta_0)\| \le \gamma} |R_T(\theta,\pi)| = o_{p\pi}(1)$ *for some nonrandom matrices* B_T *for which* $\lambda_{\min}(B_T) \to \infty$.

ASSUMPTION 3: $(B_T^{-1\prime}D\ell_T(\theta_0,\bullet), \mathcal{J}_{T\bullet}) \Rightarrow (G_\bullet, \mathcal{J}_\bullet)$ *(as processes indexed by* $\pi \in \Pi$*) for some stochastic process* $\{(G_\pi, \mathcal{J}_\pi) : \pi \in \Pi\}$ *that has bounded continuous sample paths with probability one and for which the* $s \times s$ *matrix* $\mathcal{J}_\pi$ *is symmetric* $\forall \pi \in \Pi$ *and satisfies* $0 < \inf_{\pi \in \Pi} \lambda_{\min}(\mathcal{J}_\pi) \le \sup_{\pi \in \Pi} \lambda_{\max}(\mathcal{J}_\pi) < \infty$ *with probability one*.

A useful sufficient condition for Assumption 2 is as follows.

ASSUMPTION 2*: *For all* $\gamma_T \to 0$, $\sup_{\theta \in \Theta : \|\theta-\theta_0\| \le \gamma_T} |R_T(\theta,\pi)|/(1 + \|B_T(\theta-\theta_0)\|)^2 = o_{p\pi}(1)$.

We use the Taylor expansion for functions with left/right (ℓ/r) partial derivatives developed in E1 to provide a sufficient condition for Assumption 2* that relies on smoothness of $\ell_T(\theta,\pi)$ in θ. This condition covers the two examples of this paper. The Appendix provides an additional sufficient condition for Assumption 2* that does not require smoothness of $\ell_T(\theta,\pi)$.

ASSUMPTION 2^{2^*}: (a) *For each* $\pi \in \Pi$, *the domain of* $\ell_T(\theta, \pi)$ *as a function of* θ *includes a set* Θ^+ *that satisfies* (i) $\Theta^+ - \theta_0$ *equals the intersection of a union of orthants and an open cube* $C(\mathbf{0}, \varepsilon)$ *for some* $\varepsilon > 0$ *and* (ii) $\Theta \cap S(\theta_0, \varepsilon_1) \subset \Theta^+$ *for some* $\varepsilon_1 > 0$, *where* Θ *is the parameter space*.

(b) $\ell_T(\theta, \pi)$ *has continuous* ℓ/r *partial derivatives with respect to* θ *of order two on* Θ^+ $\forall \pi \in \Pi$, $\forall T \geq 1$ *with probability one*.

(c) *For all* $\gamma_T \to 0$,

$$\sup_{\theta \in \Theta : \|\theta - \theta_0\| \leq \gamma_T} \|B_T^{-1'}\big((\partial^2/\partial\theta\,\partial\theta')\ell_T(\theta, \pi) - (\partial^2/\partial\theta\,\partial\theta')\ell_T(\theta_0, \pi)\big)B_T^{-1}\| = o_{p\pi}(1),$$

where $(\partial/\partial\theta)\ell_T(\theta, \pi)$ *and* $(\partial^2/\partial\theta\,\partial\theta')\ell_T(\theta, \pi)$ *denote the* s *vector and* $s \times s$ *matrix of* ℓ/r *partial derivatives of* $\ell_T(\theta, \pi)$ *with respect to* θ *of orders one and two respectively*.

Assumption 2^{2^*} implies Assumption 2* with $D\ell_T(\theta_0, \pi)$ and $D^2\ell_T(\theta_0, \pi)$ of (3.3) given by $(\partial/\partial\theta)\ell_T(\theta_0, \pi)$ and $(\partial^2/\partial\theta\,\partial\theta')\ell_T(\theta_0, \pi)$ respectively. The proof is analogous to that of Lemma 1 of E1.

If Assumption 2^{2^*} holds and $-B_T^{-1'}(\partial^2/\partial\theta\,\partial\theta')\ell_T(\theta_0, \pi)B_T^{-1} = \mathscr{J}_\pi + o_{p\pi}(1)$ for some nonrandom matrix $\mathscr{J}_\pi$, then Assumption 2* holds with $D\ell_T(\theta_0, \pi)$ of (3.3) given by $(\partial/\partial\theta)\ell_T(\theta_0, \pi)$ and with $D^2\ell_T(\theta_0, \pi)$ of (3.3) given by either $(\partial^2/\partial\theta\,\partial\theta')\ \ell_T(\theta_0, \pi)$ or $-B_T'\mathscr{J}_\pi B_T$.

In quasi-log likelihood cases, Assumption 3 is implied by the weak convergence of the normalized score process and Hessian indexed by $\pi \in \Pi$. This often holds by a functional central limit theorem (CLT) and a uniform law of large numbers (LLN). Thus, $G_\bullet$ is often a mean zero Gaussian process. For examples of the verification of Assumption 3, see Andrews and Ploberger (1994, 1995, 1996).

Assumption 3 allows the normalized information matrix $\mathscr{J}_{T\pi}$ to be random even in the limit as $T \to \infty$ (to cover models with stochastic trends). For models with no stochastic trends, the following is sufficient for Assumption 3.

ASSUMPTION 3*: $B_T^{-1'}D\ell_T(\theta_0, \bullet) \Rightarrow G_\bullet$ (*as a process indexed by* $\pi \in \Pi$) *for some stochastic process* $\{G_\pi : \pi \in \Pi\}$ *that has bounded continuous sample paths with probability one*. $\mathscr{J}_{T\pi}$ *is nonrandom and does not depend on* T. $\mathscr{J}_\pi$ $(:= \mathscr{J}_{T\pi})$ *is symmetric* $\forall \pi \in \Pi$, $\sup_{\pi \in \Pi} \lambda_{\max}(\mathscr{J}_\pi) < \infty$, *and* $\inf_{\pi \in \Pi} \lambda_{\min}(\mathscr{J}_\pi) > 0$.

To see the particular form the quadratic approximation of (3.3) takes for GMM and minimum distance estimators, see Section 7 of E2. For an example of a semi-parametric estimator, see Section 9 of E2.

3.4. *Asymptotic Distributions of the Unrestricted Estimator and the Objective Function*

Before obtaining the asymptotic distribution of $\hat{\theta}_\pi$, we need to establish its rate of convergence.

ASSUMPTION 4: $B_T(\hat{\theta}_\pi - \theta_0) = O_{p\pi}(1)$.

Sufficient conditions for Assumption 4 are given in the following lemma.

LEMMA 1: *Assumptions* 1, 2*, *and* 3 *imply Assumption* 4.

The proof of Lemma 1 and other results below are given in the Appendix.

Next, we consider a local approximation to the parameter space Θ after it is shifted and rescaled. The following Assumption 5 is exactly the same as in E1 and E2. Assumptions $5^*, \dots, 5^{4^*}$ stated in E1 and E2 provide primitive sufficient conditions for Assumption 5. For brevity, we only specify the simplest of these here.

ASSUMPTION 5: *For some sequence of scalar constants* $\{b_T : T \geq 1\}$ *for which* $b_T \to \infty$ *and* $b_T \leq c\lambda_{\min}(B_T)$ *for some* $0 < c < \infty$, $\{B_T(\Theta - \theta_0)/b_T : T \geq 1\}$ *is locally approximated by a cone* Λ.

See E1 or E2 for the definition of "locally approximated by a cone."

The following sufficient condition for Assumption 5 covers the two examples of this paper. We say that a set $\Gamma \subset R^s$ is *locally equal* to a set $\Lambda \subset R^s$ if $\Gamma \cap C(\mathbf{0}, \varepsilon) = \Lambda \cap C(\mathbf{0}, \varepsilon)$ for some $\varepsilon > 0$.

ASSUMPTION 5*: (a) $\Theta - \theta_0$ *is locally equal to a cone* $\Lambda \subset R^s$.
(b) $B_T = b_T I_s$ *for some scalar constants* $\{b_T : T \geq 1\}$ *for which* $b_T \to \infty$.

For each $\pi \in \Pi$, define the random variable $\hat{\lambda}_{T\pi}$ as follows: $\hat{\lambda}_{T\pi} \in \mathrm{cl}(\Lambda)$ and

$$q_T\left(\hat{\lambda}_{T\pi}, \pi\right) = \inf_{\lambda \in \Lambda} q_T(\lambda, \pi). \tag{3.4}$$

When the cone Λ of Assumption 5 is convex, $\hat{\lambda}_{T\pi}$ is uniquely defined and the normalized estimator $B_T(\hat{\theta}_\pi - \theta_0)$ is asymptotically equivalent to $\hat{\lambda}_{T\pi}$ and has an asymptotic distribution.

ASSUMPTION 6: Λ *is convex*.

The asymptotic distribution of $B_T(\hat{\theta}_\pi - \theta_0)$ is given by that of $\hat{\lambda}_\pi$. By definition, $\hat{\lambda}_\pi \in \mathrm{cl}(\Lambda)$ and

$$q\left(\hat{\lambda}_\pi, \pi\right) = \inf_{\lambda \in \Lambda} q(\lambda, \pi) \qquad \forall \pi \in \Pi, \quad \text{where} \tag{3.5}$$
$$q(\lambda, \pi) := (\lambda - Z_\pi)' \mathcal{J}_\pi (\lambda - Z_\pi) \qquad \text{and} \qquad Z_\pi := \mathcal{J}_\pi^{-1} G_\pi.$$

Under Assumption 6, $\hat{\lambda}_\pi$ is uniquely defined.

THEOREM 1: (a) *Suppose Assumptions* 2–6 *hold*. *Then*, $B_T(\hat{\theta}_\pi - \theta_0) = \hat{\lambda}_{T\pi} + o_{p\pi}(1)$.

(b) *Suppose Assumptions* 2–6 *hold*. *Then* $\hat{\lambda}_{T\bullet} \Rightarrow \hat{\lambda}_\bullet$ *and* $B_T(\hat{\theta}_\bullet - \theta_0) \Rightarrow \hat{\lambda}_\bullet$.

(c) *Suppose Assumptions* 2–5 *hold*. *Then*,

$$\ell_T\big(\hat{\theta}_\bullet, \bullet\big) - \ell_T(\theta_0, \bullet) \Rightarrow \frac{1}{2}\Big(Z'_\bullet \mathscr{I}_\bullet Z_\bullet - \inf_{\lambda \in \Lambda} q(\lambda, \bullet)\Big) = \frac{1}{2}\hat{\lambda}'_\bullet \mathscr{I}_\bullet \hat{\lambda}_\bullet \quad \text{and}$$

$$\sup_{\pi \in \Pi}\Big(\ell_T\big(\hat{\theta}_\pi, \pi\big) - \ell_T(\theta_0, \pi)\Big) \rightarrow_d \frac{1}{2}\sup_{\pi \in \Pi}\Big(Z'_\pi \mathscr{I}_\pi Z_\pi - \inf_{\lambda \in \Lambda} q(\lambda, \pi)\Big)$$

$$= \frac{1}{2}\sup_{\pi \in \Pi} \hat{\lambda}'_\pi \mathscr{I}_\pi \hat{\lambda}_\pi.$$

COMMENT: Theorem 1(b) is used below to determine the asymptotic distribution of a Wald test statistic. Theorem 1(c) is used below to determine the asymptotic distribution of a QLR test statistic.

3.5. *Asymptotic Distributions of Subvectors of the Unrestricted Estimator*

We now provide the asymptotic distribution of subvectors of $B_T(\hat{\theta}_\pi - \theta_0)$ by partitioning θ as in E1 and E2 and by partitioning $\hat{\theta}_\pi$, θ_0, B_T, G_π, $\mathscr{I}_\pi$, Z_π, and $\hat{\lambda}_\pi$ conformably with θ:

$$\theta = \begin{pmatrix} \theta_* \\ \psi \end{pmatrix} = \begin{pmatrix} \beta \\ \delta \\ \psi \end{pmatrix}, \qquad \hat{\theta}_\pi = \begin{pmatrix} \hat{\theta}_{*\pi} \\ \hat{\psi}_\pi \end{pmatrix} = \begin{pmatrix} \hat{\beta}_\pi \\ \hat{\delta}_\pi \\ \hat{\psi}_\pi \end{pmatrix}, \qquad \theta_0 = \begin{pmatrix} \theta_{*0} \\ \psi_0 \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \delta_0 \\ \psi_0 \end{pmatrix}, \tag{3.6}$$

$$B_T = \begin{bmatrix} B_{*T} & B_{*\psi T} \\ B_{\psi * T} & B_{\psi T} \end{bmatrix} = \begin{bmatrix} B_{\beta T} & B_{\beta\delta T} & B_{\beta\psi T} \\ B_{\delta\beta T} & B_{\delta T} & B_{\delta\psi T} \\ B_{\psi\beta T} & B_{\psi\delta T} & B_{\psi T} \end{bmatrix},$$

$$G_\pi = \begin{pmatrix} G_{*\pi} \\ G_{\psi\pi} \end{pmatrix} = \begin{pmatrix} G_{\beta\pi} \\ G_{\delta\pi} \\ G_{\psi\pi} \end{pmatrix}, \qquad \mathscr{I}_\pi = \begin{bmatrix} \mathscr{I}_{*\pi} & \mathscr{I}_{*\psi\pi} \\ \mathscr{I}_{\psi * T} & \mathscr{I}_{\psi\pi} \end{bmatrix} = \begin{bmatrix} \mathscr{I}_{\beta\pi} & \mathscr{I}_{\beta\delta\pi} & \mathscr{I}_{\beta\psi\pi} \\ \mathscr{I}_{\delta\beta\pi} & \mathscr{I}_{\delta\pi} & \mathscr{I}_{\delta\psi\pi} \\ \mathscr{I}_{\psi\beta\pi} & \mathscr{I}_{\psi\delta\pi} & \mathscr{I}_{\psi\pi} \end{bmatrix},$$

$$Z_\pi = \begin{pmatrix} Z_{*\pi} \\ Z_{\psi\pi} \end{pmatrix} = \begin{pmatrix} Z_{\beta\pi} \\ Z_{\delta\pi} \\ Z_{\psi\pi} \end{pmatrix}, \qquad \text{and} \qquad \hat{\lambda}_\pi = \begin{pmatrix} \hat{\lambda}_{*\pi} \\ \hat{\lambda}_{\psi\pi} \end{pmatrix} = \begin{pmatrix} \hat{\lambda}_{\beta\pi} \\ \hat{\lambda}_{\delta\pi} \\ \hat{\lambda}_{\psi\pi} \end{pmatrix},$$

where $\theta_* \in R^{p+q}$, $\beta \in R^p$, $\delta \in R^q$, and $\psi \in R^r$. We further partition β, $\hat{\beta}_\pi$, $\hat{\lambda}_{\beta\pi}$, and $Z_{\beta\pi}$ into

$$(3.7) \qquad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \quad \hat{\beta}_\pi = \begin{pmatrix} \hat{\beta}_{1\pi} \\ \hat{\beta}_{2\pi} \end{pmatrix}, \quad \hat{\lambda}_{\beta\pi} = \begin{pmatrix} \hat{\lambda}_{\beta_1\pi} \\ \hat{\lambda}_{\beta_2\pi} \end{pmatrix}, \quad \text{and} \quad Z_{\beta\pi} = \begin{pmatrix} Z_{\beta_1\pi} \\ Z_{\beta_2\pi} \end{pmatrix},$$

where $\beta_1 \in R^{p_1}$, $\beta_2 \in R^{p_2}$, and $p_1 + p_2 = p$. Let $B_{\beta_1 T}$ denote the upper $p_1 \times p_1$ block of $B_{\beta T}$. Let $B_{\beta_1\delta T}$ and $B_{\beta_1\psi T}$ denote the upper p_1 rows of $B_{\beta\delta T}$ and $B_{\beta\psi T}$ respectively.

The subvectors of θ are categorized as follows. The vector β_1 consists of the parameters that are restricted by the null hypothesis. The vector β_2 consists of nuisance parameters that lie on the boundary of the parameter space. The vector δ consists of nuisance parameters that do not lie on the boundary of the parameter space. The vector ψ consists of nuisance parameters that satisfy a block diagonality condition with respect to the other parameters. They may or may not lie on the boundary of the parameter space.

The defining features of the parameters ψ and δ are the following.

ASSUMPTION 7: (a) *$\mathcal{J}_\pi$ is block diagonal between θ_* and ψ $\forall \pi \in \Pi$. That is, $\mathcal{J}_{*\psi\pi} = \mathcal{J}'_{\psi * T} = \mathbf{0}$ $\forall \pi \in \Pi$.*

(b) *The set Λ of Assumption 5 is a product set $\Lambda_\beta \times \Lambda_\delta \times \Lambda_\psi$, where $\Lambda_\beta \subset R^p$, $\Lambda_\delta \subset R^q$, and $\Lambda_\psi \subset R^r$ are cones.*

ASSUMPTION 8: $\Lambda_\delta = R^q$.

Under Assumption 7,

$$(3.8) \qquad Z_{*\pi} = \mathcal{J}_{*\pi}^{-1} G_{*\pi}, \qquad Z_{\psi\pi} = \mathcal{J}_{\psi\pi}^{-1} G_{\psi\pi}, \text{ and } Z_{\beta\pi} = H Z_{*\pi}, \text{ where}$$
$$H := \left[I_p \vdots \mathbf{0} \right] \in R^{p \times (p+q)}.$$

For $\lambda = (\lambda'_\beta, \lambda'_\delta, \lambda'_\psi)' \in R^s$, we define

$$(3.9) \qquad q_\beta(\lambda_\beta, \pi) = (\lambda_\beta - Z_{\beta\pi})'(H \mathcal{J}_{*\pi}^{-1} H')^{-1} (\lambda_\beta - Z_{\beta\pi}) \qquad \text{and}$$
$$q_\psi(\lambda_\psi, \pi) = (\lambda_\psi - Z_{\psi\pi})' \mathcal{J}_{\psi\pi} (\lambda_\psi - Z_{\psi\pi}).$$

The asymptotic distributions of subvectors of $B_T(\hat{\theta}_\pi - \theta_0)$ are as follows.

THEOREM 2: (a) *Suppose Assumptions 2–8 hold. Then, $B_{\beta T}(\hat{\beta}_\bullet - \beta_0) \Rightarrow \hat{\lambda}_{\beta\bullet}$ provided $B_{\beta\delta T} = \mathbf{0}$ and $B_{\beta\psi T} = \mathbf{0}$, where $\hat{\lambda}_{\beta\pi}$ solves $q_\beta(\hat{\lambda}_{\beta\pi}) = \inf_{\lambda_\beta \in \Lambda_\beta} q_\beta(\lambda_\beta, \pi)$; $B_{\delta T}(\hat{\delta}_\bullet - \delta_0) \Rightarrow \mathcal{J}_{\delta\bullet}^{-1} G_{\delta\bullet} - \mathcal{J}_{\delta\bullet}^{-1} \mathcal{J}_{\delta\beta\bullet} \hat{\lambda}_{\beta\bullet}$ provided $B_{\delta\beta T} = \mathbf{0}$ and $B_{\delta\psi T} = \mathbf{0}$; $B_{\psi T}(\hat{\psi}_\bullet - \psi_0) \Rightarrow \hat{\lambda}_{\psi\bullet}$ provided $B_{\psi\beta T} = \mathbf{0}$ and $B_{\psi\delta T} = \mathbf{0}$, where $\hat{\lambda}_{\psi\pi}$ solves $q_\psi(\hat{\lambda}_{\psi\pi}, \pi) = \inf_{\lambda_\psi \in \Lambda_\psi} q_\psi(\lambda_\psi, \pi)$; and the convergence of these three terms holds jointly.*

(b) *Suppose Assumptions* 2–5, 7, *and* 8 *hold*. *Then*

$$\ell_T(\hat{\theta}_\bullet,\bullet)-\ell_T(\theta_0,\bullet)$$
$$\Rightarrow \frac{1}{2}\left(Z'_{\beta\bullet}(H\mathscr{I}_{*\bullet}^{-1}H')^{-1}Z_{\beta\bullet}-\inf_{\lambda_\beta\in\Lambda_\beta}q_\beta(\lambda_\beta,\bullet)\right)$$
$$+\frac{1}{2}G'_{\delta\bullet}\mathscr{I}_{\delta\bullet}^{-1}G_{\delta\bullet}+\frac{1}{2}\left(Z'_{\psi\bullet}\mathscr{I}_{\psi\bullet}Z_{\psi\bullet}-\inf_{\lambda_\psi\in\Lambda_\psi}q_\psi(\lambda_\psi,\bullet)\right) \quad \textit{and}$$

$$\sup_{\pi\in\Pi}\left(\ell_T(\hat{\theta}_\pi,\pi)-\ell_T(\theta_0,\pi)\right)$$
$$\to^d \frac{1}{2}\sup_{\pi\in\Pi}\Big(Z'_{\beta\pi}(H\mathscr{I}_{*\pi}^{-1}H')^{-1}Z_{\beta\pi}-\inf_{\lambda_\beta\in\Lambda_\beta}q_\beta(\lambda_\beta,\pi)$$
$$+G'_{\delta\pi}\mathscr{I}_{\delta\pi}^{-1}G_{\delta\pi}+Z'_{\psi\pi}\mathscr{I}_{\psi\pi}Z_{\psi\pi}-\inf_{\lambda_\psi\in\Lambda_\psi}q_\psi(\lambda_\psi,\pi)\Big)$$
$$=\frac{1}{2}\sup_{\pi\in\Pi}\left(\hat{\lambda}'_{\beta\pi}(H\mathscr{I}_{*\pi}^{-1}H')^{-1}\hat{\lambda}_{\beta\pi}+G'_{\delta\pi}\mathscr{I}_{\delta\pi}^{-1}G_{\delta\pi}+\hat{\lambda}'_{\psi\pi}\mathscr{I}_{\psi\pi}\hat{\lambda}_{\psi\pi}\right).$$

When Λ_β is defined by equality and/or inequality constraints, which is often the case, a closed form expression is available for $\hat{\lambda}_{\beta\pi}$. Theorem 5, (6.5), and (6.6) of E1 give the appropriate expression when a subscript π is added to Z_β, $\hat{j}$, CF_j, $P_{L(j)}$, and A. As an example, suppose $\Lambda_\beta = R^+$. Then,

$$\hat{\lambda}_{\beta\pi}=\max\{Z_{\beta\pi},0\}. \tag{3.10}$$

4. EXAMPLES (CONTINUED)

4.1. *GARCH Example*

We consider the Gaussian QML estimator of θ. The Gaussian quasi-log likelihood function is

$$\ell_T(\theta,\pi):=-\frac{T}{2}\ln(2\tilde{\pi})-\frac{1}{2}\sum_{t=1}^{T}\ln h_t^*(\theta,\pi)-\frac{1}{2}\sum_{t=1}^{T}e_t^2(\theta)/h_t^*(\theta,\pi), \tag{4.1}$$

where $\tilde{\pi}=3.14\ldots$ denotes the number pi. Assumption 1* is verified in the Appendix.

Note that $h_t^*(\theta_0,\pi)=\delta_0+\pi^{t-1}(h_1^*(\theta_0,\pi)-\delta_0)$. In consequence, when $h_1^*(\theta_0,\pi)=\delta_0$, we have $h_t^*(\theta_0,\pi)=\delta_0$ for all t and $\ell_T(\theta_0,\pi)$ does not depend on π. If $h_1^*(\theta_0,\pi)\neq\delta_0$, then $\ell_T(\theta_0,\pi)$ depends on π through the terms $\pi^{t-1}(h_1^*(\theta_0,\pi)-\delta_0)$.

Next, we define the components of the quadratic approximation of $\ell_T(\theta, \pi)$. Let

(4.2) $$h_t(\theta, \pi) := \delta + \beta_1 \sum_{k=0}^{\infty} \pi^k e_{t-k-1}^2(\theta) \qquad \text{and}$$

$$\ell_{tt}(\theta, \pi) := -\frac{1}{2}\ln(2\tilde{\pi}) - \frac{1}{2}\ln(h_t(\theta, \pi)) - \frac{1}{2}e_t^2(\theta)/h_t(\theta, \pi).$$

(The double subscript on $\ell_{tt}(\theta, \pi)$ is used to distinguish $\ell_{tt}(\theta, \pi)$ from $\ell_T(\theta, \pi)$ when $t = T$.) Note that $h_t(\theta, \pi)$ is the unobserved conditional variance given the parameters (θ, π) with the initial condition $h_1^*(\theta, \pi)$ replaced by an infinite weighted sum of lagged values of $e_t^2(\theta)$. Also, $\ell_{tt}(\theta, \pi)$ is the corresponding unobserved tth contribution to the quasi-log likelihood. The asymptotic behavior of the actual quasi-log likelihood formed using $h_t^*(\theta, \pi)$ is shown to be equivalent to that based on $h_t(\theta, \pi)$.

The components of the quadratic approximation of $\ell_T(\theta, \pi)$ at θ_0 are:

(4.3) $$D\ell_T(\theta_0, \pi) := \sum_{t=1}^{T} \frac{\partial}{\partial \theta}\ell_{tt}(\theta_0, \pi), \qquad D^2\ell_T(\theta_0, \pi) := -T\mathscr{J}_\pi,$$

$$B_T := T^{1/2} I_{r+2},$$

$$\frac{\partial}{\partial \theta}\ell_{tt}(\theta_0, \pi) = \left(\frac{1}{2}(z_t^2 - 1) \sum_{k=0}^{\infty} \pi^k z_{t-k-1}^2, \frac{1}{2\delta_0}(z_t^2 - 1), z_t X_t'/\delta_0^{1/2} \right)',$$

and

$$\mathscr{J}_\pi := -E\frac{\partial^2}{\partial \theta \, \partial \theta'}\ell_{tt}(\theta_0, \pi)$$

$$= \frac{1}{2} \begin{pmatrix} \dfrac{2c}{1-\pi^2} + \dfrac{1}{(1-\pi)^2} & \dfrac{1}{\delta_0(1-\pi)} & 0 \\ \dfrac{1}{\delta_0(1-\pi)} & \delta_0^{-2} & 0 \\ 0 & 0 & 2\delta_0^{-1} E X_t X_t' \end{pmatrix},$$

where

$$c := (Ez_t^4 - 1)/2.$$

Assumptions 2^{2^*} and 3* are verified in the Appendix. The verification of Assumption 3* uses the central limit theorem for square integrable, stationary and ergodic, martingale difference sequences applied to $\{(\partial \ell_{tt}(\theta_0, \pi_1)/\partial \theta', \ldots, \partial \ell_{tt}(\theta_0, \pi_J)/\theta')' : t = \ldots, 0, 1, \ldots\}$ for arbitrary $(\pi_1, \ldots, \pi_J)$ to obtain the convergence in distribution of the finite dimensional distributions of the process $B_T^{-1'} D\ell_T(\theta, \bullet)$. In consequence, the limit process $\{G_\pi : \pi \in \Pi\}$ of Assumption 3* is a mean zero Gaussian process with covariance function $\mathrm{Cov}(G_{\pi_1}, G_{\pi_2}) = \mathscr{I}_{\pi_1,\pi_2} := (E(\partial \ell_{tt}(\theta_0, \pi_1)/\partial\theta)(\partial \ell_{tt}(\theta_0, \pi_2)/\partial\theta')'$. Some calculations show that

$$(4.4)\qquad \mathscr{I}_{\pi_1,\pi_2} := \begin{pmatrix} \frac{c}{2}\left(\frac{2c}{1-\pi_1\pi_2} + \frac{1}{(1-\pi_1)(1-\pi_2)}\right) & \frac{c}{2\delta_0(1-\pi_2)} & \frac{1}{2\delta_0^{1/2}}\sum_{k=0}^{\infty}\pi_2^k E z_t^3 z_{t-k-1}^2 X_t' \\ \frac{c}{2\delta_0(1-\pi_1)} & \frac{c}{2\delta_0^2} & \frac{1}{2\delta_0^{3/2}} E z_t^3 X_t' \\ \frac{1}{2\delta_0^{1/2}}\sum_{k=0}^{\infty}\pi_1^k E z_t^3 z_{t-k-1}^2 X_t & \frac{1}{2\delta_0^{3/2}} E z_t^3 X_t & \delta_0^{-1} E z_t^2 X_t X_t' \end{pmatrix}.$$

Assumption 4 holds by Lemma 1. Assumptions 5* and 6 hold with $\Lambda := R^+ \times R^{r+1}$.

There is no parameter β_2 in this example, so $\beta = \beta_1$. The vector θ is partitioned as $\theta = (\beta_1, \delta, \psi')'$ and $\theta_* := (\beta_1, \delta)'$. Assumption 7(a) holds, because $\mathscr{I}_\pi$ is block diagonal by (4.3). Assumption 7(b) holds, because Λ is a product set with

$$(4.5)\qquad \Lambda_\beta := \Lambda_{\beta_1} := R^+, \qquad \Lambda_\delta := R, \qquad \text{and} \qquad \Lambda_\psi := R^r.$$

Theorem 2 and (3.10) provide the asymptotic distributions of $T^{1/2}\hat\beta_{1\bullet}, T^{1/2}(\hat\delta_\bullet - \delta_0), T^{1/2}(\hat\psi_\bullet - \psi_0)$, and $\sup_{\pi\in\Pi}(\ell_T(\hat\theta_\pi,\pi) - \ell_T(\theta_0,\pi))$ for this example, where $\hat\beta_{1\bullet} = \hat\beta_\bullet$, because all the requisite conditions have been verified and B_T is diagonal. We have

$$(4.6)\qquad T^{1/2}\hat\beta_{1\bullet} \Rightarrow \hat\lambda_{\beta_1\bullet}, \qquad \text{where}$$
$$\hat\lambda_{\beta_1\pi} = \max\{Z_{\beta_1\pi}, 0\} \qquad \text{and}$$
$$Z_{\beta_1\pi} := H\mathscr{I}_{*\pi}^{-1}G_{*\pi} \sim N(0, H\mathscr{I}_{*\pi}^{-1}\mathscr{I}_{*\pi,\pi}\mathscr{I}_{*\pi}^{-1}H') = N(0, 1-\pi^2).$$

Here $H = (1,0)$ and $\mathscr{I}_{*\pi,\pi}$ denotes the upper 2×2 block of $\mathscr{I}_{\pi,\pi}$. Note that $\hat\lambda_{\beta_1\pi}$ has a half-normal distribution.

The process $Z_{\beta_1\pi}$ is a Gaussian process with covariance function that is fairly simple. Equations (4.3) and (4.4) and some calculations show that

$$(4.7)\qquad \mathrm{cov}(Z_{\beta_1\pi_1}, Z_{\beta_1\pi_2}) = H\mathscr{I}_{*\pi_1}^{-1}\mathscr{I}_{*\pi_1,\pi_2}\mathscr{I}_{*\pi_2}^{-1}H' = \frac{(1-\pi_1^2)(1-\pi_2^2)}{1-\pi_1\pi_2}.$$

Let $\{\tilde Z_i : i \geq 1\}$ be a sequence of iid standard normal random variables. Simple calculations show that $(1-\pi^2)\sum_{i=0}^{\infty}\pi^i\tilde Z_i$ is a Gaussian process that has the same covariance function as $Z_{\beta_1\pi}$. Thus,

$$(4.8)\qquad Z_{\beta_1\pi} \simeq (1-\pi^2)\sum_{i=0}^{\infty}\pi^i\tilde Z_i \quad \text{and} \quad \hat\lambda_{\beta_1\pi} \simeq \max\left\{(1-\pi^2)\sum_{i=0}^{\infty}\pi^i\tilde Z_i, 0\right\},$$

where $\simeq$ denotes equality in distribution of the stochastic processes indexed by $\pi \in \Pi$. One can simulate the processes $Z_{\beta_1\pi}$ and $\hat{\lambda}_{\beta_1\pi}$ easily by simulating the truncated process $(1-\pi^2)\sum_{i=0}^{TR}\pi^i\tilde{Z}_i$ for some TR large.

By Theorem 2(a), $T^{1/2}(\hat{\delta}_\bullet - \delta_0) \Rightarrow \hat{\lambda}_{\delta\bullet}$, where $\hat{\lambda}_{\delta\pi} = 2\delta_0^2 G_\delta - \delta_0(1-\pi)^{-1}\max\{Z_{\beta_1\pi}, 0\}$, $G_\delta := G_{\delta\pi}$ does not depend on π (because it is a mean zero Gaussian process whose covariance function is given by the (2, 2) element of $\mathscr{I}_{\pi_1,\pi_2}$ in (4.4), which is independent of π), and $G_{*\pi} := (G_{\beta_1\pi}, G_\delta)' \sim N(\mathbf{0}, \mathscr{I}_{*\pi,\pi})$.

Note that $G_\psi := G_{\psi\pi}$ and $\mathscr{J}_\psi := \mathscr{J}_{\psi\pi}$ do not depend on π. By Theorem 2(a) and the fact that $\Lambda_\psi = R^r$, $B_{\psi\beta T} = \mathbf{0}$, and $B_{\psi\delta T} = \mathbf{0}$, we have $T^{1/2}(\hat{\psi}_\bullet - \psi_0) \Rightarrow \hat{\lambda}_\psi$, where $\hat{\lambda}_\psi = Z_\psi := \mathscr{J}_\psi^{-1} G_\psi \sim N(\mathbf{0}, \mathscr{J}_\psi^{-1})$ and $\mathscr{J}_\psi^{-1} = \delta_0(EX_tX_t')^{-1}$. Let $\hat{\pi}$ denote any estimator of π. Then, $T^{1/2}(\hat{\psi}_{\hat{\pi}} - \psi_0) \Rightarrow \hat{\lambda}_\psi \sim N(\mathbf{0}, (EW_tW_t'/\text{var}_t(\theta_0))^{-1})$. This implies that the extremum estimator of ψ from the maximization of $\ell_T(\theta, \pi)$ over $(\theta, \pi) \in \Theta \times \Pi$ is asymptotically normally distributed when the GARCH-MA parameter is zero just as it is when the GARCH-MA parameter is positive. (The preceding result for $\hat{\psi}_{\hat{\pi}}$ holds because $\inf_{\pi\in\Pi} T^{1/2}(\hat{\psi}_\pi - \psi_0) \le T^{1/2}(\hat{\psi}_{\hat{\pi}} - \psi_0) \le \sup_{\pi\in\Pi} T^{1/2}(\hat{\psi}_\pi - \psi_0)$ and the lower and upper bounds both have asymptotic distribution given by $\hat{\lambda}_\psi$ by the continuous mapping theorem.)

By Theorem 2(b), (4.6), (4.7), and $\mathscr{J}_{\delta\pi} = \delta_0^{-2}/2$, we obtain

$$\text{(4.9)} \qquad \sup_{\pi\in\Pi}\left(\ell_T\left(\hat{\theta}_\pi, \pi\right) - \ell_T(\theta_0, \pi)\right) \to^d \frac{1}{2}\sup_{\pi\in\Pi}\left(\max{}^2\{Z_{\beta_1\pi}, 0\}c/(1-\pi^2)\right) + G_\delta^2\delta_0^2 + \frac{1}{2}\hat{\lambda}_\psi'\mathscr{J}_\psi\hat{\lambda}_\psi.$$

4.2. *Random Coefficient Example*

We consider the Gaussian QML estimator, which is based on the assumption that ε_t and η_t are normally distributed and independent of X_t. The Gaussian quasi-log likelihood function is

$$\text{(4.10)} \qquad \ell_T(\theta, \pi) := -\frac{T}{2}\ln(2\tilde{\pi}) - \frac{1}{2}\sum_{t=1}^T \ln(\delta_2 + X_t'\Omega(\beta_1, \delta_1, \pi)X_t) - \frac{1}{2}\sum_{t=1}^T (Y_t - \psi_2 - X_t'\psi_1)^2/(\delta_2 + X_t'\Omega(\beta_1, \delta_1, \pi)X_t).$$

Assumption 1* is verified in the Appendix.

The quadratic approximation of $\ell_T(\theta, \pi)$ at θ_0 is defined as follows. Let

$$\text{(4.11)} \qquad \begin{aligned} &W_t := (X_t', 1)', \qquad \tilde{W}_t(\pi) := (X_{1t}^*(\pi), X_{2t}^*(\delta_{10})', 1)', \\ &X_{1t}^*(\pi) := X_{1t}'\Omega_1(\pi)X_{1t}, \qquad X_{2t}^*(\delta_{10}) := (\partial/\partial\delta_1)X_{2t}'\Omega_2(\delta_{10})X_{2t}, \\ &\text{res}_t(\theta) := Y_t - \psi_2 - X_t'\psi_1, \qquad \text{and} \\ &\text{var}_t(\theta, \pi) := \delta_2 + X_t'\Omega(\beta_1, \delta_1, \pi)X_t. \end{aligned}$$

Note that $\text{var}_t(\theta, \pi)$ does not depend on π when $\theta = \theta_0$, because $X_t'\Omega(\beta_1, \delta_1, \pi)X_t = \beta_1 X_{1t}'\Omega_1(\pi)X_{1t} + X_{2t}'\Omega_2(\delta_1)X_{2t}$. In consequence, we denote $\text{var}_t(\theta_0, \pi)$ by $\text{var}_t(\theta_0)$.

Define

$$
\text{(4.12)} \quad D\ell_T(\theta_0 l, \pi) := \sum_{t=1}^{T} \left(\frac{\text{res}_t^2(\theta_0) - \text{var}_t(\theta_0)}{2\text{var}_t^2(\theta_0)} \tilde{W}_t(\pi)', \frac{\text{res}_t(\theta_0)}{\text{var}_t(\theta_0)} W_t' \right)',
$$

$$
D^2\ell_T(\theta_0, \pi) := -T\mathcal{J}_\pi,
$$

$$
\mathcal{J}_{T\pi} := \mathcal{J}_\pi := \begin{bmatrix} \frac{1}{2} E\tilde{W}_t(\pi)\tilde{W}_t(\pi)'/\text{var}_t^2(\theta_0) & \mathbf{0} \\ \mathbf{0} & EW_tW_t'/\text{var}_t(\theta_0) \end{bmatrix},
$$

$$
B_T := T^{1/2} I_s, \quad \text{and}
$$

$$
Z_{T\pi} := \mathcal{J}_\pi^{-1} T^{-1/2} D\ell_T(\theta_0, \pi).
$$

It is shown in the Appendix that the quadratic approximation of (3.3) holds (in particular, Assumption 2^{2^*} holds) under the assumptions above and the moment conditions below.

We assume that

$$
\text{(4.13)} \quad E\|\varepsilon_t X_t\|^4 < \infty, \qquad E\|\eta_t\|^4 \|X_t\|^8 < \infty,
$$

$$
\text{(4.14)} \quad EW_tW_t' > 0, \qquad E(\varepsilon_t^2 - 1)^2 \tilde{W}_t(\pi)\tilde{W}_t(\pi)' > 0, \qquad \forall \pi \in \Pi, \quad \text{and}
$$

$$
E\tilde{X}_t\tilde{X}_t' > 0,
$$

where $\tilde{X}_t$ is a vector that includes the constant 1 and $X_{tj}X_{tk}$ for all $j = 1, \ldots, k$ and $k = 1, \ldots, b + c$ and "> 0" denotes "is positive definite."

Assumption 3* is verified in the Appendix. The verification uses the CLT for iid square integrable random vectors to obtain the convergence in distribution of the finite dimensional distributions of the process $T^{-1/2}D\ell_T(\theta_0, \bullet)$. In consequence, the limit process $\{G_\pi : \pi \in \Pi\}$ is a mean zero Gaussian process with covariance function $\text{cov}(G_{\pi_1}, G_{\pi_2}) = \mathcal{J}_{\pi_1, \pi_2} := ED\ell_T(\theta_0, \pi_1) D\ell_T(\theta_0, \pi_2)'$, where

$$
\text{(4.15)} \quad \mathcal{J}_{\pi_1, \pi_2} := \begin{bmatrix} \frac{1}{4} E \frac{(\text{res}_t^2(\theta_0) - \text{var}_t(\theta_0))^2}{\text{var}_t^4(\theta_0)} \tilde{W}_t(\pi_1)\tilde{W}_t(\pi_2)' & \frac{1}{2} E \frac{\text{res}_t^3(\theta_0)}{\text{var}_t^3(\theta_0)} \tilde{W}_t(\pi_1) W_t' \\ \frac{1}{2} E \frac{\text{res}_t^3(\theta_0)}{\text{var}_t^3(\theta_0)} W_t \tilde{W}_t(\pi_2)' & EW_tW_t'/\text{var}_t(\theta_0) \end{bmatrix}.
$$

Assumption 4 holds by Lemma 1. Assumptions 5* and 6 hold with $\Lambda := R^+ \times R^{s-1}$.

We partition θ as follows: $\beta := \beta_1$, $\delta := (\delta_1', \delta_2)'$, $\psi := (\psi_1', \psi_2)'$, and $\theta_* := (\beta_1', \delta_1', \delta_2)'$. With this partitioning, Assumptions 7 and 8 hold. In particular, by (4.12), $\mathcal{J}_\pi$ is block diagonal between θ_* and ψ. The set Λ is a product set

$\Lambda_\beta \times \Lambda_\delta \times \Lambda_\psi$ with

$$\Lambda_\beta := \Lambda_{\beta_1} := R^+, \qquad \Lambda_\delta := R^{g+1}, \qquad \text{and} \qquad \Lambda_\psi := R^{b+c+1}. \tag{4.16}$$

Theorem 2 provides the asymptotic distributions of $T^{1/2}\hat{\beta}_{1\bullet}, T^{1/2}(\hat{\delta}_\bullet - \delta_0), T^{1/2}(\hat{\psi}_\bullet - \delta_0), T^{1/2}(\hat{\psi}_\bullet - \psi_0)$, and $\sup_{\pi \in \Pi}(\ell_T(\hat{\theta}_\pi, \pi) - \ell_T(\theta_0, \pi))$, where $\hat{\beta}_{1\bullet} = \hat{\beta}_\bullet$, because all of the requisite conditions have been verified and B_T is diagonal. In particular,

$$T^{1/2}\hat{\beta}_{1\bullet} \Rightarrow \hat{\lambda}_{\beta_1\bullet} = \max\{Z_{\beta_1\bullet}, 0\}, \qquad \text{where} \tag{4.17}$$

$$Z_{\beta_1\pi} = H\mathscr{J}_{*\pi}^{-1}G_{*\pi} \sim N(0, H\mathscr{J}_{*\pi}^{-1}\mathscr{I}_{*\pi,\pi}\mathscr{J}_{*\pi}^{-1}H'),$$

using (3.10). Thus, $\hat{\lambda}_{\beta_1\pi}$ has a half-normal distribution. Unlike the GARCH Example, the covariance function of $\hat{\lambda}_{\beta_1\bullet}$ does not simplify. It is given by $\text{cov}(\hat{\lambda}_{\beta_1\pi_1}, \hat{\lambda}_{\beta_1\pi_2}) = H\mathscr{J}_{*\pi_1}^{-1}\mathscr{I}_{*\pi_1,\pi_2}\mathscr{J}_{*\pi_2}^{-1}H'$.

By Theorem 2(a), $T^{1/2}(\hat{\delta}_\bullet - \delta_0) \Rightarrow \hat{\lambda}_{\delta\bullet}$, where $\hat{\lambda}_{\delta\pi} = \mathscr{J}_\delta^{-1}G_\delta - \mathscr{J}_\delta^{-1}\mathscr{J}_{\delta\beta_1\pi} \times \max\{Z_{\beta_1\pi}, 0\}$, $G_{*\pi} := (G_{\beta_1\pi}, G_\delta)' \sim N(\mathbf{0}, \mathscr{I}_{*\pi,\pi})$, and G_δ and $\mathscr{J}_\delta := \mathscr{J}_{\delta\pi}$ do not depend on π.

By Theorem 2(a) and $\Lambda_\psi = R^{b+c+1}$, we obtain $\hat{\lambda}_\psi = Z_\psi := \mathscr{J}_\psi^{-1}G_\psi \sim N(\mathbf{0}, \mathscr{J}_\psi^{-1})$ and $T^{1/2}(\hat{\psi}_\bullet - \psi_0) \Rightarrow \hat{\lambda}_\psi \sim N(\mathbf{0}, (EW_tW_t'/\text{var}_t(\theta_0))^{-1})$, where $\hat{\lambda}_\psi, Z_\psi, G_\psi := G_{\psi\pi}$, and $\mathscr{J}_\psi := \mathscr{J}_{\psi\pi}$ do not depend on π. Let $\hat{\pi}$ denote any estimator of π. Then, $T^{1/2}(\hat{\psi}_{\hat{\pi}} - \psi_0) \Rightarrow \hat{\lambda}_\psi \sim N(\mathbf{0}, (EW_tW_t'/\text{var}_t(\theta_0))^{-1})$, for the same reason as in the GARCH Example. This implies that the extremum estimator of ψ from the maximization of $\ell_T(\theta, \pi)$ over $(\theta, \pi) \in \Theta \times \Pi$ has an asymptotic normal distribution.

By Theorem 2(b), we have

$$\sup_{\pi \in \Pi}\left(\ell_T\left(\hat{\theta}_\pi, \pi\right) - \ell_T(\theta_0, \pi)\right) \to^d \frac{1}{2}\sup_{\pi \in \Pi}\max{}^2\{Z_{\beta\pi}, 0\}/(H\mathscr{J}_{*\pi}^{-1}H') + G_\delta'\mathscr{J}_\delta^{-1}G_\delta + \hat{\lambda}_\psi'\mathscr{J}_\psi\hat{\lambda}_\psi. \tag{4.18}$$

5. THE QUASI-LIKELIHOOD RATIO TEST

5.1. *The QLR Test Statistic*

In this section, we define the QLR test statistic and give the asymptotic distribution of the QLR statistic under the null hypothesis.

The null hypothesis is

$$H_0 : \beta_1 = \beta_{1*}, \tag{5.1}$$

for some specified vector $\beta_{1*} \in R^{p_1}$. The form of the null hypothesis is built into part (a) of the following assumption.

ASSUMPTION 9: (a) *For some* $\beta_{1*} \in R^{p_1}$, $\Theta_0 = \{\theta \in \Theta : \theta = (\beta_{1*}', \beta_2', \delta', \psi')'$ *for some* $\beta_2 \in R^{p_2}$, $\delta \in R^q$, *and* $\psi \in R^r\}$.

(b) $B_{\beta_1 \delta T} = \mathbf{0}$ *and* $B_{\beta_1 \psi T} = \mathbf{0}$.

(c) Θ *is a product set with respect to* $(\beta_1', \beta_2', \delta', \psi')'$ *local to* θ, $\forall \theta \in \Theta_0$. *That is,* $\Theta \cap S(\theta, \varepsilon) = (\mathscr{B}_1 \times \mathscr{B}_2 \times \Delta \times \Psi) \cap S(\theta, \varepsilon)$ *for some* $\mathscr{B}_1 \subset R^{p_1}$, $\mathscr{B}_2 \subset R^{p_2}$, $\Delta \subset R^q$, $\Psi \subset R^r$, *and* $\varepsilon > 0$, $\forall \theta \in \Theta_0$.

(d) $\Lambda_\beta = \Lambda_{\beta_1} \times \Lambda_{\beta_2}$ *for some cones* $\Lambda_{\beta_1} \subset R^{p_1}$ *and* $\Lambda_{\beta_2} \subset R^{p_2}$.

As shown below, in conjunction with Assumptions 5, 7, and 8, Assumption 9 implies that Assumption 5 holds with Θ replaced by Θ_0 and Λ replaced by

$$(5.2) \qquad \Lambda_0 := \{\mathbf{0}\} \times \Lambda_{\beta_2} \times R^q \times \Lambda_\psi,$$

where Λ_ψ and Λ_{β_2} are the same as in Assumptions 7 and 9 respectively.

The *restricted* (by H_0) extremum estimator of θ_0 for given π is denoted $\hat{\theta}_{0\pi}$. By definition, $\theta_{0\pi} \in \Theta_0$ $\forall \pi \in \Pi$ and

$$(5.3) \qquad \ell_T\left(\hat{\theta}_{0\pi}, \pi\right) = \sup_{\theta \in \Theta_0} \ell_T(\theta, \pi) + o_{p\pi}(1).$$

We partition $\hat{\theta}_{0\pi}$ as $(\beta_{1*}', \hat{\beta}_{20\pi}', \hat{\delta}_{0\pi}', \hat{\psi}_{0\pi}')'$ conformably with θ.

The QLR statistic is defined by

$$(5.4) \qquad QLR_T := -2\left(\sup_{\pi \in \Pi} \ell_T\left(\hat{\theta}_{0\pi}, \pi\right) - \sup_{\pi \in \Pi} \ell_T\left(\hat{\theta}_\pi, \pi\right)\right) + o_p(1).$$

5.2. *Results for the Restricted Parameter Space* Θ_0

Here we determine the asymptotic distribution of the maximum of the estimator objective function over the restricted parameter space Θ_0. We do so by applying the results of Section 3 with Θ replaced by the null parameter space Θ_0 throughout.

We use the following restricted analogues of Assumptions 1 and 4:

ASSUMPTION 1_0: $\hat{\theta}_{0\pi} = \theta_0 + o_{p\pi}(1)$.

ASSUMPTION 4_0: $B_T(\hat{\theta}_{0\pi} - \theta_0) = O_{p\pi}(1)$.

We note that Assumption 1* is a sufficient condition for Assumption 1_0 provided $\theta_0 \in \Theta_0$, which we assume here, because $\theta_0 \in \Theta_0 \subset \Theta$. Sufficient conditions for Assumption 4_0 are given in Theorem 3 below.

Let $\hat{\lambda}_{0\pi}$ be a minimizer of $q(\lambda, \pi)$ over $\mathrm{cl}(\Lambda_0)$. That is, $\hat{\lambda}_{0\pi} \in \mathrm{cl}(\Lambda_0)$ $\forall \pi \in \Pi$ and

$$(5.5) \qquad q\left(\hat{\lambda}_{0\pi}, \pi\right) = \inf_{\lambda \in \Lambda_0} q(\lambda, \pi).$$

We partition $\hat{\lambda}_{0\pi}$ conformably with θ:

$$\hat{\lambda}_{0\pi} = \left(\hat{\lambda}'_{\beta 0\pi}, \hat{\lambda}'_{\delta 0\pi}, \hat{\lambda}'_{\psi 0\pi} \right)' = \left(\mathbf{0}', \hat{\lambda}'_{\beta_2 0\pi}, \hat{\lambda}'_{\delta 0\pi}, \hat{\lambda}'_{\psi 0\pi} \right)'. \tag{5.6}$$

THEOREM 3: *Suppose* $\theta_0 \in \Theta_0$.

(a) *Suppose Assumptions* 5 *and* 7–9 *hold. Then, Assumption* 5 *holds with* Θ *and* Λ *replaced by* Θ_0 *and* Λ_0, *respectively, where* Λ_ψ *and* Λ_{β_2} *are the same as in Assumptions* 7 *and* 9.

(b) *Assumptions* 1_0, 2^*, *and* 3 *imply Assumption* 4_0.

(c) *Suppose Assumptions* 2, 3, *and* 4_0 *hold. Then,* $\ell_T(\hat{\theta}_{0\pi}, \pi) = \ell_T(\theta_0, \pi) + \frac{1}{2} Z'_{T\pi} \mathscr{J}_{T\pi} Z_{T\pi} - \frac{1}{2} \inf_{\theta \in \Theta_0} q_T(B_T(\theta - \theta_0), \pi) + o_{p\pi}(1)$.

(d) *Suppose Assumptions* 2, 3, 4_0, 5, *and* 7–9 *hold. Then,*

$$\ell_T\left(\hat{\theta}_{0\bullet}, \bullet\right) - \ell_T(\theta_0, \bullet) \Rightarrow \frac{1}{2}\left(Z'_\bullet \mathscr{J}_\bullet Z_\bullet - \inf_{\lambda \in \Lambda_0} q(\lambda, \bullet) \right) = \frac{1}{2} \hat{\lambda}'_{0\bullet} \mathscr{J}_\bullet \hat{\lambda}_{0\bullet}$$

and

$$\sup_{\pi \in \Pi} \left(\ell_T\left(\hat{\theta}_{0\pi}, \pi\right) - \ell_T(\theta_0, \pi) \right) \rightarrow^d \frac{1}{2} \sup_{\pi \in \Pi} \left(Z'_\pi \mathscr{J}_\pi Z_\pi - \inf_{\lambda \in \Lambda_0} q(\lambda, \pi) \right)$$
$$= \frac{1}{2} \sup_{\pi \in \Pi} \hat{\lambda}'_{0\pi} \mathscr{J}_\pi \hat{\lambda}_{0\pi}.$$

(e) *Suppose Assumptions* 3, 5, *and* 7–9 *hold. Then,* $Z'_\pi \mathscr{J}_\pi Z_\pi - \inf_{\lambda \in \Lambda_0} q(\lambda, \pi) = Z'_{\beta\pi}(H \mathscr{J}_{*\pi}^{-1} H')^{-1} Z_{\beta\pi} - \inf_{\lambda_\beta \in \{\mathbf{0}\} \times \Lambda_{\beta_2}} q_\beta(\lambda_\beta, \pi) + G'_{\delta\pi} \mathscr{J}_{\delta\pi}^{-1} G_{\delta\pi} + Z'_{\psi\pi} \mathscr{J}_{\psi\pi} Z_{\psi\pi} - \inf_{\lambda_\psi \in \Lambda_\psi} q_\psi(\lambda_\psi, \pi) = \hat{\lambda}'_{\beta 0\pi}(H \mathscr{J}^{-1}_{*\pi} H')^{-1} \hat{\lambda}_{\beta 0\pi} + G'_{\delta\pi} \mathscr{J}_{\delta\pi}^{-1} G_{\delta\pi} + \hat{\lambda}'_{\psi\pi} \mathscr{J}_{\psi\pi} \hat{\lambda}_{\psi\pi}$.

COMMENT: The convergence in part (d) is joint with that of Theorems 1 and 2 because all of the results follow from an application of the continuous mapping theorem to the process $(B_T^{-1'} D\ell_T(\theta_0, \bullet), \mathscr{J}_{T\bullet})$.

5.3. *The Asymptotic Null Distribution of the QLR Statistic*

The testing applications that we consider for the QLR test are ones for which $\ell_T(\theta_0, \pi)$ does not depend on π. In particular, dependence of $\ell_T(\theta_0, \pi)$ on π through initial conditions in a time series context is not permitted. Furthermore, we require that $G_{\delta\pi}$, $G_{\psi\pi}$, $\mathscr{J}_{\delta\pi}$, and $\mathscr{J}_{\psi\pi}$ do not depend on π.

ASSUMPTION 10: (a) $\ell_T(\theta_0, \pi)$ *does not depend on* π *for all* T *large and is denoted* $\ell_T(\theta_0)$.

(b) $G_{\delta\pi}$, $G_{\psi\pi}$, $\mathscr{J}_{\delta\pi}$, *and* $\mathscr{J}_{\psi\pi}$ *do not depend on* π *and are denoted* G_δ, G_ψ, $\mathscr{J}_\delta$, *and* $\mathscr{J}_\psi$ *respectively.*

Assumption 10(a) is violated in some time series models if the initial conditions are chosen inappropriately. For example, in the GARCH Example, if

$h_1^*(\theta, \pi)$ is such that $h_1^*(\theta_0, \pi) \neq \delta_0$, then $\ell_T(\theta_0, \pi)$ depends on π by an $o_{p\pi}(1)$ term. The reason Assumption 10(b) typically holds is that $\ell_T(\theta, \pi)$ does not depend on π for all $\theta \in \Theta_0$ and Θ_0 is determined by the subvector β_1, not by δ or ψ. Thus, starting at $\theta_0 \in \Theta_0$, a change in δ or ψ still leaves θ_0 in Θ_0 and still leaves $\ell_T(\theta_0, \pi)$ independent of π. In consequence, $D_\delta \ell_T(\theta_0, \pi)$, $D_\psi \ell_T(\theta_0, \pi)$, $\mathscr{J}_{\delta T\pi}$, $\mathscr{J}_{\psi T\pi}$, and the (normalized) limits of these terms, $G_{\delta\pi}$, $G_{\psi\pi}$, $\mathscr{J}_{\delta\pi}$, and $\mathscr{J}_{\psi\pi}$, do not depend on π, where $D\ell_T(\theta_0, \pi) := (D_\beta \ell_T(\theta_0, \pi)', D_\delta \ell_T(\theta_0, \pi)', D_\psi \ell_T(\theta_0, \pi)')'$.

Note the implications of Assumption 10. First, by part (a), $QLR_T = \sup_{\pi \in \Pi}(\ell_T(\hat{\theta}_\pi, \pi) - \ell_T(\theta_0, \pi)) - \sup_{\pi \in \Pi}(\ell_T(\hat{\theta}_{0\pi}, \pi) - \ell_T(\theta_0, \pi)) + o_p(1)$ and the asymptotic null distributions of the first two summands are given in Theorems 2(b) and 3(d) respectively. Second, by Assumption 10(b), many of the limit random variables in Theorems 2(b) and 3(e) do not depend on π and can be pulled out of the expressions involving the supremum over $\pi \in \Pi$. Furthermore, the terms that do not depend on π are exactly the same in the limit expressions for $\sup_{\pi \in \Pi}(\ell_T(\hat{\theta}_\pi, \pi) - \ell_T(\theta_0, \pi))$ and $\sup_{\pi \in \Pi}(\ell_T(\hat{\theta}_{0\pi}, \pi) - \ell_T(\theta_0, \pi))$. Thus, they cancel when one considers the asymptotic distribution of QLR_T for $\theta_0 \in \Theta_0$.

We now state the asymptotic distribution of QLR_T for parameter values $\theta_0 \in \Theta_0$.

THEOREM 4: *Suppose* $\theta_0 \in \Theta_0$ *and Assumptions* 2–5, 4_0, *and* 7–10 *hold. Then*,

(a) $QLR_T = \sup_{\pi \in \Pi}(Z'_{T\pi}\mathscr{J}_{T\pi}Z_{T\pi} - \inf_{\lambda \in \Lambda} q_T(\lambda, \pi)) - \sup_{\pi \in \Pi}(Z'_{T\pi}\mathscr{J}_{T\pi}Z_{T\pi} - \inf_{\lambda \in \Lambda_0} q_T(\lambda, \pi)) + o_p(1)$,

(b) $QLR_T \to_d AD_{QLR} := \sup_{\pi \in \Pi}(Z'_\pi \mathscr{J}_\pi Z_\pi - \inf_{\lambda \in \Lambda} q(\lambda, \pi)) - \sup_{\pi \in \Pi}(Z'_\pi \cdot \mathscr{J}_\pi Z_\pi - \inf_{\lambda \in \Lambda_0} q(\lambda, \pi)) = \sup_{\pi \in \Pi} \hat{\lambda}'_\pi \mathscr{J}_\pi \hat{\lambda}_\pi - \sup_{\pi \in \Pi} \hat{\lambda}'_{0\pi} \mathscr{J}_\pi \hat{\lambda}_{0\pi} \geq 0$, *and*

(c) $AD_{QLR} = \sup_{\pi \in \Pi}(Z'_{\beta\pi}(H\mathscr{J}_{*\pi}^{-1}H')^{-1}Z_{\beta\pi} - \inf_{\lambda_\beta \in \Lambda_\beta} q_\beta(\lambda_\beta, \pi)) - \sup_{\pi \in \Pi}(Z'_{\beta\pi}(H\mathscr{J}_{*\pi}^{-1}H')^{-1}Z_{\beta\pi} - \inf_{\lambda_\beta \in \{0\} \times \Lambda_{\beta_2}} q_\beta(\lambda_\beta, \pi)) = \sup_{\pi \in \Pi} \hat{\lambda}'_{\beta\pi} \cdot (H\mathscr{J}_{*\pi}^{-1}H')^{-1}\hat{\lambda}_{\beta\pi} - \sup_{\pi \in \Pi} \hat{\lambda}'_{\beta 0\pi}(H\mathscr{J}_{*\pi}^{-1}H')^{-1}\hat{\lambda}_{\beta 0\pi} \geq 0$.

COMMENTS: 1. AD_{QLR} stands for "asymptotic distribution of the QLR statistic."

2. The expression for AD_{QLR} in Theorem 4(c) has the advantage over the expression in 4(b) that the dimension p of β is often much smaller than the dimension s of θ.

3. A general method of obtaining critical values corresponding to AD_{QLR} is by simulation. (This is also true for the *RQLR*, Wald, and score tests considered below.) If Λ_β is defined by inequality constraints, which is often the case, then $\hat{\lambda}_{\beta\pi}$ is the solution to a quadratic programming problem. In consequence, $\hat{\lambda}_{\beta\pi}$ can be simulated very quickly. Programs for solving quadratic programming problems are available in GAUSS and Matlab. When unknown nuisance parameters appear in the definition of AD_{QLR}, they can be replaced by consistent estimates in order to carry out the simulations.

The asymptotic null distribution of QLR_T, AD_{QLR}, given in Theorem 4(b)–(c) simplifies in various cases. First, if the estimator objective function does not depend on π for any $\theta \in \Theta$ (or, equivalently, if Π contains a single element), then the suprema over $\pi \in \Pi$ disappear everywhere, the $Z'_\pi \mathscr{J}_\pi Z_\pi$ terms cancel in Theorem 4(b), and the $Z'_{\beta\pi}(H\mathscr{J}_{*\pi}^{-1}H')^{-1}Z_{\beta\pi}$ terms cancel in Theorem 4(c).

Second, if no parameter β_2 appears, as occurs in the GARCH and Random Coefficient Examples, then the terms $\sup_{\pi \in \Pi}(Z'_{\beta\pi}(H\mathscr{J}_{*\pi}^{-1}H')^{-1}Z_{\beta\pi} - \inf_{\lambda_\beta \in \{\mathbf{0}\} \times \Lambda_{\beta_2}} q_\beta(\lambda_\beta, \pi))$ and $\sup_{\pi \in \Pi} \hat{\lambda}'_{\beta 0\pi}(H\mathscr{J}_{*\pi}^{-1}H')^{-1}\hat{\lambda}_{\beta 0\pi}$ in Theorem 4(c) are both zero.

Third, if $\Lambda_\beta = R^p$, then $\inf_{\lambda_\beta \in \Lambda_\beta} q_\beta(\lambda_\beta, \pi) = 0$, $\hat{\lambda}_{\beta\pi} = Z_{\beta\pi}$, without loss of generality (wlog) no parameter β_2 appears (because it can be absorbed into δ), the terms in the previous paragraph are zero, and

$$(5.7) \qquad AD_{QLR} = \sup_{\pi \in \Pi} Z'_{\beta\pi}(H\mathscr{J}_{*\pi}^{-1}H')^{-1} Z_{\beta\pi}.$$

This corresponds to the classical case of an unrestricted alternative hypothesis and no nuisance parameters on the boundary of the parameter space.

Fourth, suppose the following assumption holds:

ASSUMPTION 11: (a) *$G_{*\pi} \sim N(\mu_\pi, \mathscr{I}_{*\pi})$ conditional on some σ-field $\mathscr{F}$, for some (possibly random) $(p+q) \times (p+q)$ matrix-valued process $\mathscr{I}_{*\pi}$ and $p+q$ vector-valued process μ_π that are $\mathscr{F}$ measurable and whose sample paths are bounded and continuous functions of π on Π with probability one.*

(b) *$\mathscr{I}_{*\pi} = c_\pi \mathscr{J}_{*\pi}$ for some (possibly random) scalar c_π with* $\inf_{\pi \in \Pi} c_\pi > 0$.

Typically $\mu_\pi = \mathbf{0}$ when θ_0 is in the null hypothesis, as is considered here. We allow for $\mu_\pi \neq \mathbf{0}$, because this assumption also is used below for local power results and in such cases one usually has $\mu_\pi \neq \mathbf{0}$.

Assumption 11(a) often holds as a result of a central limit theorem. If Assumption 11(a) holds and $\ell_T(\theta, \pi)$ is a correctly specified log-likelihood function, then the information matrix equality for any given π implies that Assumption 11(b) holds with $c_\pi = 1$. Assumption 11(b) holds for LS estimators of regression models with $c_\pi = \sigma^2$ provided Assumption 11(a) holds and the regression errors are homoskedastic conditional on the regressors with variance σ^2. Assumption 11(b) holds for GMM and minimum distance estimators with $c_\pi = 1$ provided an asymptotically optimal weight matrix is employed.

When Assumption 11 holds with $\mu_\pi = \mathbf{0}$ and $\Lambda_\beta = R^p$, we have

$$(5.8) \qquad Z'_{\beta\pi}(H\mathscr{J}_{*\pi}^{-1}H')^{-1} Z_{\beta\pi} \sim c_\pi \chi^2_p \qquad \forall \pi \in \Pi,$$

where χ^2_p denotes a random variable with a chi-squared distribution with p degrees of freedom. Thus, AD_{QLR} is the supremum of a (rescaled) chi-squared process.

When Assumption 11 holds with $\mu_\pi = \mathbf{0}$, $\Lambda_\beta \neq R^p$, and Λ_β is a convex set, the distribution of $\hat{\lambda}'_{\beta\pi}(H\mathscr{J}_{*\pi}^{-1}H')^{-1}\hat{\lambda}_{\beta\pi}$ for fixed π is that of c_π times a mixture of chi-square random variables. This follows from Theorem 3.1 of Shapiro (1985). See Shapiro (1985, Sec. 4) for formulae for the mixing weights when $p \leq 4$.

Given that Assumption 11(a) holds, the additional Assumption 11(b) reduces the number of nuisance parameters that appears in the asymptotic distribution AD_{LR}. To see this, let A_π be a $p \times p$ (possibly random) matrix that is symmetric and nonsingular with probability one for each $\pi \in \Pi$. Our leading choice for A_π is

$$A_\pi = \text{diag}^{1/2}(H \mathscr{J}_{*\pi}^{-1} H'). \tag{5.9}$$

Let $\Lambda_{\beta A \pi} := A_\pi^{-1} \Lambda_\beta$ and $Z_{\beta A \pi} := A_\pi^{-1} Z_{\beta \pi}$. Define $\hat{\lambda}_{\beta A \pi}$ such that $\hat{\lambda}_{\beta A \pi} \in \text{cl}(\Lambda_{\beta A \pi})$ and

$$q_{\beta A}\left(\hat{\lambda}_{\beta A \pi}, \pi\right) = \inf_{\lambda_\beta \in \Lambda_{\beta A \pi}} q_{\beta A}(\lambda_\beta, \pi), \qquad \text{where}$$

$$q_{\beta A}(\lambda_\beta, \pi) := (\lambda_\beta - Z_{\beta A \pi})' (A_\pi^{-1} H \mathscr{J}_{*\pi}^{-1} H' A_\pi^{-1})^{-1} (\lambda_\beta - Z_{\beta A \pi}) \quad \text{for} \quad \lambda_\beta \in R^p. \tag{5.10}$$

Define $\hat{\lambda}_{\beta 0 A \pi}$ as $\hat{\lambda}_{\beta A \pi}$ is defined but with $\Lambda_{\beta A \pi}$ replaced by $\Lambda_{\beta 0 A \pi} := A_\pi^{-1}(\{\mathbf{0}\} \times \Lambda_{\beta_2})$.

LEMMA 2: *For any $p \times p$ (possibly random) matrix A_π that is symmetric and nonsingular for each $\pi \in \Pi$ with probability one, $\hat{\lambda}_{\beta \pi} = A_\pi \hat{\lambda}_{\beta A \pi}$ and $\hat{\lambda}_{\beta 0 \pi} = A_\pi \hat{\lambda}_{\beta 0 A \pi}$.*

COMMENTS: 1. By Lemma 2, $\hat{\lambda}_{\beta\pi}' (H \mathscr{J}_{*\pi}^{-1} H')^{-1} \hat{\lambda}_{\beta\pi}$ equals $\hat{\lambda}_{\beta A \pi}' \cdot (A_\pi^{-1} H \mathscr{J}_{*\pi}^{-1} H' A_\pi^{-1})^{-1} \hat{\lambda}_{\beta A \pi}$ and analogously with $\hat{\lambda}_{\beta\pi}$ and $\hat{\lambda}_{\beta A \pi}$ replaced by $\hat{\lambda}_{\beta 0 \pi}$ and $\hat{\lambda}_{\beta 0 A \pi}$. Under Assumption 11(a), the distributions of these terms depend on the nuisance parameters (or nuisance random variables) in $A_\pi^{-1} H \mathscr{J}_{*\pi}^{-1} H' A_\pi^{-1}$ and in the (conditional) covariance matrix of $Z_{\beta A \pi}$, viz., $A_\pi^{-1} H \mathscr{J}_{*\pi}^{-1} \mathscr{I}_{*\pi} \mathscr{J}_{*\pi}^{-1} H' A_\pi^{-1}$. Take A_π as in (5.9) and suppose Assumption 11(b) also holds. Then, the matrix $A_\pi^{-1} H \mathscr{J}_{*\pi}^{-1} \mathscr{I}_{*\pi} \mathscr{J}_{*\pi}^{-1} H' A_\pi^{-1}$ equals $c_\pi A_\pi^{-1} H \mathscr{J}_{*\pi}^{-1} H' A_\pi^{-1}$ and knowledge of the former implies knowledge of $A_\pi^{-1} H \mathscr{J}_{*\pi}^{-1} H' A_\pi^{-1}$ up to scale. In this case, the total number of nuisance parameters in these two matrices reduces from p^2 to $1 + p(p-1)/2$ when c_π is unknown and from p^2 to $p(p-1)/2$ when c_π is known.

2. The proof of Lemma 2 follows easily from the fact that $q_\beta(\lambda_\beta, \pi) = q_{\beta A}(A_\pi^{-1} \lambda_\beta, \pi)$.

5.4. *The Rescaled Quasi-Likelihood Ratio Test*

In this section, we introduce a *rescaled* QLR statistic that eliminates, or reduces the number of, nuisance parameters that appear in the asymptotic null distribution of the QLR statistic. We consider the common case where Assumption 11(b) holds, but $\mathscr{I}_{*\pi} \neq \mathscr{J}_{*\pi}$. For example, the latter often occurs in likelihood scenarios with model misspecification, such as the GARCH Example when the innovations have a nonnormal distribution. In such cases, the asymptotic distribution of the QLR statistic depends on some nuisance parameters.

Let $\hat{c}_\pi$ be an estimator of c_π (defined in Assumption 11(b)). We append the following assumption to Assumption 11:

ASSUMPTION 11: (c) $\hat{c}_\bullet \Rightarrow c_\bullet$ *jointly with the convergence of Assumption* 3.

When c_π is nonrandom, Assumption 11(c) holds if $\hat{c}_\pi = c_\pi + o_{p\pi}(1)$.

The *rescaled* QLR statistic is defined to be

$$(5.11)\qquad RQLR_T := -2\Big(\sup_{\pi\in\Pi}\ell_T\big(\hat{\theta}_{0\pi},\pi\big)/\hat{c}_\pi - \sup_{\pi\in\Pi}\ell_T\big(\hat{\theta}_\pi,\pi\big)/\hat{c}_\pi\Big) + o_p(1).$$

The asymptotic null distribution of $RQLR_T$ is given in the following theorem.

THEOREM 5: *Suppose* $\theta_0 \in \Theta_0$ *and Assumptions* 2–5, 4_0, *and* 7–11 *hold. Then*, $RQLR_T \to^d AD_{RQLR} := \sup_{\pi\in\Pi} \hat{\lambda}'_{\beta\pi}(H\mathcal{J}_{*\pi}^{-1}\mathcal{I}_{*\pi}\mathcal{J}_{*\pi}^{-1}H')^{-1}\hat{\lambda}_{\beta\pi} - \sup_{\pi\in\Pi}\hat{\lambda}'_{\beta 0\pi}(H\mathcal{J}_{*\pi}^{-1}\mathcal{I}_{*\pi}\mathcal{J}_{*\pi}^{-1}H')^{-1}\hat{\lambda}_{\beta 0\pi} \geq 0$.

COMMENT: By Assumption 11(a), the weight matrix $(H\mathcal{J}_{*\pi}^{-1}\mathcal{I}_{*\pi}\mathcal{J}_{*\pi}^{-1}H')^{-1}$ on the right-hand side of Theorem 5 equals the inverse of the covariance matrix of $Z_{\beta\pi}$ conditional on $(\mathcal{J}_{*\pi}, \mathcal{I}_{*\pi})$. Let $A_\pi = \text{diag}^{1/2}(H\mathcal{J}_{*\pi}^{-1}\mathcal{I}_{*\pi}\mathcal{J}_{*\pi}^{-1}H')$. By Lemma 2, $\hat{\lambda}'_{\beta\pi}(H\mathcal{J}_{*\pi}^{-1}\mathcal{I}_{*\pi}\mathcal{J}_{*\pi}^{-1}H')^{-1}\hat{\lambda}_{\beta\pi}$ equals $\hat{\lambda}_{\beta A\pi}\cdot(A_\pi^{-1}H\mathcal{J}_{*\pi}^{-1}\mathcal{I}_{*\pi}\mathcal{J}_{*\pi}^{-1}H'A_\pi^{-1})^{-1}\hat{\lambda}_{\beta A\pi}$ and likewise with $\hat{\lambda}_{\beta\pi}$ and $\hat{\lambda}_{\beta A\pi}$ replaced by $\hat{\lambda}_{\beta 0\pi}$ and $\hat{\lambda}_{\beta 0 A\pi}$. If $\mu_\pi = \mathbf{0}$, the distributions of these terms depend only on the nuisance parameters in $A_\pi^{-1}H\mathcal{J}_{*\pi}^{-1}\mathcal{I}_{*\pi}\mathcal{J}_{*\pi}^{-1}H'A_\pi^{-1}$. Since the main diagonal elements are all ones, there are at most $p(p-1)/2$ unknown nuisance parameters. When $p=1$, there are no nuisance parameters and

$$(5.12)\qquad \hat{\lambda}'_{\beta\pi}(H\mathcal{J}_{*\pi}^{-1}\mathcal{I}_{*\pi}\mathcal{J}_{*\pi}^{-1}H')^{-1}\hat{\lambda}_{\beta\pi} \sim \begin{cases} N_\pi^2 & \text{when } \Lambda_\beta = R, \\ \max^2\{N_\pi, 0\} & \text{when } \Lambda_\beta = R^+, \end{cases}$$

where $N_\pi \sim N(0,1)\ \forall \pi \in \Pi$. Of course, the correlations between N_{π_1} and N_{π_2} for $\pi_1, \pi_2 \in \Pi$ might depend on nuisance parameters. When $p \geq 2$, the distribution of $\hat{\lambda}'_{\beta\pi}(H\mathcal{J}_{*\pi}^{-1}\mathcal{I}_{*\pi}\mathcal{J}_{*\pi}^{-1}H')^{-1}\hat{\lambda}_{\beta\pi}$ for fixed π is that of a mixture of chi-square random variables by Shapiro (1985, Thm. 3.1), provided Λ_β is convex. See Shapiro (1985, Sec. 4) for the mixing weights. The same holds for the term with $\hat{\lambda}_{\beta 0\pi}$ in place of $\hat{\lambda}_{\beta\pi}$.

5.2. *GARCH Example* (*Continued*)

The null and alternative hypotheses of interest are $H_0: \beta_1 = 0$ and $H_1: \beta_1 > 0$. A test of H_0 versus H_1 is a test for the presence of conditional heteroskedasticity.

Assumption 9(a) holds with $\beta_{1*} = 0$. Assumption 9(b) holds because $B_T = T^{1/2} I_{r+2}$. Assumption 9(c) holds with $\mathscr{B}_1 = R^+$, $\Delta = R$, $\Psi = R^r$, and $p_2 = 0$. Assumption 9(d) holds because $p_2 = 0$.

The restricted parameter space is $\Theta_0 := \{\theta \in \Theta : \theta = (0, \delta, \psi')'\}$. When the initial condition is $h_1^*(\theta, \pi) = \delta$, the quasi-log likelihood does not depend on π for any θ in Θ_0. This condition really is part of the specification of the null hypothesis because it only needs to hold under the null hypothesis and the desired null hypothesis is that the conditional variance is a constant δ_0. In this case, the restricted estimator $\hat{\theta}_{0\pi}$ does not depend on π and equals

$$\text{(5.13)} \qquad \hat{\theta}_{0\pi} = \left(0, \hat{\delta}_0, \hat{\psi}_0\right)', \qquad \text{where} \qquad \hat{\psi}_0 = \left(\sum_{t=1}^{T} X_t X_t'\right)^{-1} \sum_{t=1}^{T} X_t Y_t,$$

$$\hat{\delta}_0 = \sum_{t=1}^{T} \hat{e}_t^2 / T, \qquad \text{and} \qquad \hat{e}_t = Y_t - X_t' \hat{\psi}_0.$$

Whether or not $h_1^*(\theta, \pi) = \delta$, Assumption 1_0 holds, because Assumption 1* has already been verified and it is a sufficient condition for Assumption 1_0 when $\theta_0 \in \Theta_0$.

By Theorem 3(a), Assumptions 5 and 7–9 imply that Assumption 5 holds with Θ and Λ replaced by Θ_0 and Λ_0, respectively, where $\Lambda_0 := \{0\} \times R \times R^r$. Assumption 4_0 holds by Theorem 3(b) because Assumptions 1_0, 2*, and 3 have been verified.

Assumption 10(a) holds if the initial condition is such that $h_1^*(\theta_0, \pi) = \delta_0$, but not otherwise. An earlier section points out that the conditions of Assumption 10(b) hold in this example.

By Theorem 3(d) and (e) and Assumption 10(b), we have

$$\text{(5.14)} \qquad \sup_{\pi \in \Pi} \left(\ell_T\left(\hat{\theta}_{0\pi}, \pi\right) - \ell_T(\theta_0, \pi) \right) \to^d G_\delta^2 \delta_0^2 + \frac{1}{2} \hat{\lambda}_\psi' \mathscr{J}_\psi \hat{\lambda}_\psi .$$

By Theorem 4(b) and (c), (4.3), (4.4), (4.6), and (4.8), we have: under the null,

$$\text{(5.15)} \qquad QLR_T \to^d \sup_{\pi \in \Pi} \hat{\lambda}_{\beta_1 \pi}' (H \mathscr{J}_{*\pi}^{-1} H')^{-1} \hat{\lambda}_{\beta_1 \pi} = c \sup_{\pi \in \Pi} \max{}^2\{Z_{\beta_1 \pi}, 0\} / (1 - \pi^2)$$

$$\simeq c \sup_{\pi \in \Pi} \max{}^2\{v_\pi, 0\}, \qquad \text{where} \qquad v_\pi := (1 - \pi^2)^{1/2} \sum_{i=0}^{\infty} \pi^i \tilde{Z}_i .$$

Here, $\{\tilde{Z}_i : i \geq 1\}$ are iid standard normal random variables. In general, the asymptotic null distribution of QLR_T depends on the nuisance parameter $c = (Ez_t^4 - 1)/2$. If the errors are normally distributed, then $c = 1$ and this nuisance parameter disappears. For this case, we have generated the asymptotic critical values by simulation. For $\Pi = [.00, .95]$, for significance levels 10%, 5%, and 1%, the critical values are 3.06, 4.33, and 7.30 respectively. These values were obtained using 40,000 simulation repetitions and the discrete grid $\Pi = \{.00, .01, \ldots, .95\}$. The same critical values were obtained using the finer grid $\Pi = \{.000, .001, \ldots, .950\}$.

Next, we consider the $RQLR_T$ statistic. By (4.3) and (4.4), Assumptions 11(a) and (b) hold with $\mu_\pi = \mathbf{0}$ and $\mathscr{I}_{*\pi} = c\mathscr{J}_{*\pi}$. If the errors are not necessarily normally distributed, we consider two estimators of c. The first employs $\hat{\theta}_{0\pi}$. The second employs $\hat{\theta}_\pi$. Both are such that Assumption 11(c) holds. The estimators are

$$(5.16)\qquad \hat{c} := \left(\frac{1}{T}\sum_{t=1}^{T} \hat{e}_t^4/\hat{\delta}_0^2 - 1\right)/2 \qquad \text{and}$$

$$\hat{c}_\pi := \left(\frac{1}{T}\sum_{t=1}^{T} e_t^4\left(\hat{\theta}_\pi\right)/h_t^*\left(\hat{\theta}_\pi, \pi\right)^2 - 1\right)/2,$$

where $\hat{e}_t$ and $\hat{\delta}_0$ are defined in (5.13) and $e_t(\theta)$ and $h_t^*(\theta, \pi)$ are defined in (2.1). We define $RQLR_T$ as in (5.11) with $\hat{c}_\pi$ given by either of the definitions in (5.16).

By Theorem 5, $RQLR_T$ has the same asymptotic null distribution as the QLR_T statistic given in (5.15), but with $H\mathscr{J}_{*\pi}^{-1}H'$ replaced by $H\mathscr{J}_{*\pi}^{-1}\mathscr{I}_{*\pi}\mathscr{J}_{*\pi}^{-1}H'$ and $c = 1$. Thus, the $RQLR_T$ test statistic has a nuisance parameter free asymptotic null distribution. Critical values for this test statistic for arbitrary c are the same as those listed above for the QLR statistic for the special case where $c = 1$.

5.6. *Random Coefficient Example* (*Continued*)

The null and alternative hypotheses of interest are: $H_0 : \beta_1 = 0$ and $H_1 : \beta_1 > 0$. Under the null, the coefficients on the regressors X_{1t} are nonrandom. Thus, a test of H_0 versus H_1 is a test for the presence of random variation in the coefficients on X_{1t}. Assumption 9(a) holds with $\beta_{1*} = 0$. Assumption 9(b) holds because $B_T = T^{1/2}I_s$. Assumption 9(c) holds with $\mathscr{B}_1 = R^+$, $\Delta = R^{g+1}$, $\Psi = R^{b+c+1}$, and $p_2 = 0$. Assumption 9(d) holds because $p_2 = 0$.

The restricted parameter space is $\Theta_0 := \{\theta \in \Theta : \theta = (0, \delta', \psi')'\}$. The quasi-log likelihood does not depend on π for any θ in Θ_0. In consequence, the restricted estimator $\hat{\theta}_{0\pi}$ does not depend on π and is denoted $\hat{\theta}_0$. Assumption 1_0 holds, because Assumption 1* has already been verified. For the same reasons as in the GARCH Example, Assumption 4_0 holds and Assumption 5 holds with Θ and Λ replaced by Θ_0 and Λ_0, respectively, where $\Lambda_0 := \{0\} \times R^{g+1} \times R^{b+c+1}$.

By Theorem 3(d) and (e), we have

$$(5.17)\qquad \sup_{\pi\in\Pi}\left(\ell_T\left(\hat{\theta}_{0\pi}, \pi\right) - \ell_T(\theta_0, \pi)\right) \to^d \frac{1}{2}\left(G_\delta'\mathscr{J}_\delta^{-1}G_\delta + \hat{\lambda}_\psi'\mathscr{J}_\psi\hat{\lambda}_\psi\right).$$

Assumption 10 holds in this example because $\tilde{W}_t(\pi)$ depends on π only through the subvector $X_{1t}^*(\pi)$. By Theorem 4(b) and (c) and (4.17), we have: under H_0,

$$(5.18)\qquad QLR_T \to^d \sup_{\pi\in\Pi} \max{}^2\{Z_{\beta_1\pi}, 0\}/(H\mathscr{J}_{*\pi}^{-1}H') = \sup_{\pi\in\Pi} \max{}^2\{v_{\beta_1\pi}, 0\},$$

where

$$v_{\beta_1\pi} := (H\mathscr{J}_{*\pi}^{-1}H')^{-1/2} Z_{\beta_1\pi} \sim N(0, c_\pi) \qquad \text{and}$$

$$c_\pi := H\mathscr{J}_{*\pi}^{-1}\mathscr{I}_{*\pi,\pi}\mathscr{J}_{*\pi}^{-1}H' / \left(H\mathscr{J}_{*\eta}^{-1}H'\right).$$

If the errors ε_t and η_t are normally distributed and are independent of X_t, then $\mathscr{I}_{*\pi,\pi} = \mathscr{J}_{*\pi}$ and $c_\pi = 1$. In this case, the distribution of $\max^2\{v_{\beta_1\pi}, 0\}$ for π fixed is nuisance parameter free with distribution function $F(x) = 1/2 + F_{\chi_1^2}(x)/2$, where $F_{\chi_1^2}(x)$ is the distribution function of a chi-square random variable with one degree of freedom. If a single random coefficient variance is under test, i.e., $b = 1$, and the errors are normal, then no nuisance parameter π appears under the alternative and the asymptotic critical value for a QLR test of significance level α is given by the $1 - 2\alpha$ quantile of a chi-square random variable with one degree of freedom. For significance levels 10%, 5%, and 1%, the critical values are 1.642, 2.706, and 5.412.

When more than one variance is under test, one can obtain critical values and p-values by simulating $\sup_{\pi \in \Pi} \max^2\{v_{\beta_1\pi}, 0\}$ with any unknown quantities replaced by consistent estimates. This can be done as follows. Let $\{\tilde{Z}_i : i = 1, \ldots, T\}$ be iid $N(0,1)$ random variables. Let

$$\hat{v}_{\beta_1\pi} := \left(H\hat{\mathscr{J}}_{*T\pi}^{-1}H'\right)^{-1/2} H\hat{\mathscr{J}}_{*T\pi}^{-1}T^{-1/2} \times \sum_{i=1}^{T} \frac{\text{res}_i^2\left(\hat{\theta}_0\right) - \text{var}_i\left(\hat{\theta}_0\right)}{2\,\text{var}_i^2\left(\hat{\theta}_0\right)} \tilde{W}_i\left(\hat{\theta}_0\right)\tilde{Z}_i, \qquad \text{where} \tag{5.19}$$

$$\hat{\mathscr{J}}_{*T\pi} := \frac{1}{2T}\sum_{t=1}^{T} \tilde{W}_t\left(\hat{\theta}_0, \pi\right)\tilde{W}_t\left(\hat{\theta}_0, \pi\right)' / \text{var}_t^2\left(\hat{\theta}_0\right) \qquad \text{and}$$

$$\tilde{W}_t(\theta, \pi) := (X_{1t}^*(\pi)', X_{2t}^*(\delta_1)', 1)'.$$

One simulates $\hat{v}_{\beta_1\pi}$ for a finite grid Π_G of π values in Π and computes $\sup_{\pi \in \Pi_G} \max^2\{\hat{v}_{\beta_1\pi}, 0\}$. The $1 - \alpha$ quantile of R such simulations is the appropriate asymptotic critical value for a level α test. Provided the mesh size of the grid goes to zero and the number of simulation repetitions goes to infinity as $T \to \infty$, the simulated critical values yield a test with the correct asymptotic rejection rate under the null. (This follows by weak convergence of $\hat{v}_{\beta_1\bullet}$ to $v_{\beta_1\bullet}$ as $T \to \infty$ and the continuous mapping theorem, where the random variables $\{Z_i : i = 1, \ldots, T\}$ are defined on the same probability space as the original sample.) The fraction of simulated values of $\sup_{\pi \in \Pi} \max^2\{\hat{v}_{\beta_1\pi}, 0\}$ that exceed the observed value of QLR_T is the simulated p-value. If necessary, for ease of computation, one can replace the sum over $i = 1, \ldots, T$ by a sum over $i = 1, \ldots, T_1$ in the definition of $\hat{v}_{\beta_1\pi}$, where $T_1 < T$. Provided $T_1 \to \infty$ as $T \to \infty$, the resulting simulated critical values and p-values are still asymptotically correct. The above simulation method is quite similar to that employed in Hansen (1996).

To avoid the computational burden of simulating $\hat{v}_{\beta_1\pi}$ for a very fine grid of π values, we recommend using a relatively course grid and defining the test statistic QLR_T with the same grid Π_G as used for the simulations.

To illustrate computational costs, we simulated asymptotic critical values for the case where two random coefficient variances are under test, one correlation parameter π_1 appears, the sample size is 100, a grid of 21 equally spaced values of π_1 in $(-1,1)$ is used, and a grid of 20 equally spaced values of π_2 on the quarter unit circle are used. The computation time for 1000 simulation repetitions was 4.9 minutes using a Pentium II 333 Mhz PC. The computation time was found to increase linearly in the number of simulation repetitions and the number of π values.

Next, we consider the $RQLR_T$ test. Assumption 11(a) holds with $\mu_\pi = 0$ and $\mathscr{I}_{*\pi} = \mathscr{I}_{*\pi,\pi}$. Assumption 11(b) holds with c_π defined above. Note that $\inf_{\pi\in\Pi} c_\pi > 0$ because $\mathscr{J}_{*\pi}$ and $\mathscr{I}_{*\pi,\pi}$, defined in (4.12) and (4.15), are positive definite by (4.14) and continuous in π and Π is compact. In this case, we define

$$(5.20)\qquad \hat{c}_\pi := H\hat{\mathscr{J}}_{*T\pi}^{-1}\hat{\mathscr{I}}_{*T\pi}\hat{\mathscr{J}}_{*T\pi}^{-1}H'/\left(H\hat{\mathscr{J}}_{*T\pi}^{-1}H'\right), \qquad \text{where}$$

$$\hat{\mathscr{I}}_{*T\pi} := \frac{1}{4T}\sum_{t=1}^{T}\frac{\left(\text{res}_t^2\left(\hat{\theta}_0\right) - \text{var}_t\left(\hat{\theta}_0\right)\right)^2}{\text{var}_t^4\left(\hat{\theta}_0\right)}\tilde{W}_t\left(\tilde{\theta}_0,\pi\right)\tilde{W}_t\left(\tilde{\theta}_0,\pi\right)'$$

and $\hat{\mathscr{J}}_{*T\pi}$ is defined in (5.19). Alternatively, we can define $\hat{c}_\pi$ with $\hat{\theta}_0$ replaced by $\hat{\theta}_\pi$ in (5.20). In either case, Assumption 11(c) holds.

By Theorem 5, the asymptotic null distribution of $RQLR_T$ is given by that of

$$(5.21)\qquad \sup_{\pi\in\Pi}\max{}^2\{\eta_{\beta_1\pi},0\}, \qquad \text{where}$$

$$\eta_{\beta_1\pi} := (H\mathscr{J}_{*\pi}^{-1}\mathscr{I}_{*\pi}\mathscr{J}_{*\pi}^{-1}H')^{-1/2}Z_{\beta_1\pi} \sim N(0,1).$$

When a single random coefficient variance is under test, the limit random variable has distribution function $F(x) = 1/2 + F_{\chi_1^2}(x)/2$ (whether or not the errors are normally distributed). Critical values are given above. When more than one parameter is under test, then simulation methods, as discussed above, can be used to obtain critical values and p-values.

6. THE WALD TEST

6.1. *Definition of the Wald Statistic*

In this section, we consider a Wald test of $H_0: \beta_1 = \beta_{1*}$ and determine its asymptotic null distribution. The test statistic defined by Wald (1943) is a quadratic form in the difference between an unrestricted QML estimator $\hat{\beta}_1$ and the value β_{1*}. We consider such a statistic in a more general context in which the unrestricted estimator $\hat{\beta}_{1\pi}$ satisfies restrictions imposed by the maintained hypothesis and is allowed to depend on a parameter π.

The resulting generalized Wald test is asymptotically equivalent to the QLR test under correct model specification in likelihood scenarios in some cases and can be made to have improved asymptotic properties under model misspecification by judicious choice of its weight matrix.

The weight matrix for the quadratic form in $\hat{\beta}_{1\pi} - \beta_{1*}$ is denoted $\hat{V}_{T\pi}^{-1}$. Conditions that it must satisfy are given below. The *Wald test statistic*, W_T, is defined as follows:

$$(6.1) \qquad W_T := \sup_{\pi \in \Pi} W_T(\pi) + o_p(1), \qquad \text{where}$$

$$W_T(\pi) := \left(\hat{\beta}_{1\pi} - \beta_{1*} \right)' B'_{\beta_1 T} \hat{V}_{T\pi}^{-1} B_{\beta_1 T} \left(\hat{\beta}_{1\pi} - \beta_{1*} \right).$$

The weight matrix $\hat{V}_{T\pi}^{-1}$ is assumed to satisfy the following assumption.

ASSUMPTION 12: (a) $\hat{V}_{T\bullet} \Rightarrow V_\bullet$ *for some stochastic process* $\{V_\pi : \pi \in \Pi\}$ *whose sample paths are bounded and continuous with probability one and the convergence is joint with* $(B_T^{-1'} D\ell_T(\theta_0, \bullet), \mathscr{J}_{T\bullet}) \Rightarrow (G_\bullet, \mathscr{J}_\bullet)$ *of Assumption* 3.
(b) $\inf_{\pi \in \Pi} \lambda_{\min}(V_\pi) > 0$ *a.s.*

We now partition the sample size T quantities $\mathscr{J}_{T\pi}$ and $\hat{\lambda}_{T\pi}$ defined in (3.3) and (3.4) conformally with θ:

$$(6.2) \qquad \mathscr{J}_{T\pi} = \begin{bmatrix} \mathscr{J}_{*T\pi} & \mathscr{J}_{*\psi T\pi} \\ \mathscr{J}_{\psi * T\pi} & \mathscr{J}_{\psi T\pi} \end{bmatrix} \quad \text{and} \quad \hat{\lambda}_{T\pi} = \left(\hat{\lambda}'_{\beta_1 T\pi}, \hat{\lambda}'_{\beta_2 T\pi}, \hat{\lambda}'_{\delta T\pi}, \hat{\lambda}'_{\psi T\pi} \right)'.$$

The following is a sufficient condition for Assumption 12 that is applicable except in unit root cases.

ASSUMPTION 12*: (a) *Assumption* 11(*a*) *holds*.
(b) $\hat{V}_{T\pi} = H_1 \hat{\mathscr{J}}_{*T\pi}^{-1} \hat{\mathscr{I}}_{*T\pi} \hat{\mathscr{J}}_{*T\pi}^{-1} H'_1$, *where* $H_1 = [I_{p_1} \vdots \mathbf{0}] \in R^{p_1 \times (p+q)}$.
(c) $\hat{\mathscr{J}}_{*T\pi} = \mathscr{J}_{*T\pi} + o_{p\pi}(1)$.
(d) $\hat{\mathscr{I}}_{*T\bullet} \Rightarrow \mathscr{I}_{*\bullet}$ *jointly with* $(B_T^{-1'} D\ell_T(\theta_0, \bullet), \mathscr{J}_{T\bullet}) \Rightarrow (G_\bullet, \mathscr{J}_\bullet)$ *of Assumption* 3.
(e) $0 < \inf_{\pi \in \Pi} \lambda_{\min}(\mathscr{I}_{*\pi}) \leq \sup_{\pi \in \Pi} \lambda_{\max}(\mathscr{I}_{*\pi}) < \infty$ *with probability one*.

When $(\mathscr{J}_{*\pi}, \mathscr{I}_{*\pi})$ are nonrandom, Assumptions 12*(c) and 12*(d) hold if $(\hat{\mathscr{J}}_{*T\pi}, \hat{\mathscr{I}}_{*T\pi}) = (\mathscr{J}_{*\pi}, \mathscr{I}_{*\pi}) + o_{p\pi}(1)$. Under Assumption 12*, $V_\pi = H_1 \mathscr{J}_{*\pi}^{-1} \mathscr{I}_{*\pi} \mathscr{J}_{*\pi}^{-1} H'_1$. When Assumption 11(b) holds, the latter simplifies to $V_\pi = c_\pi H_1 \mathscr{J}_{*\pi}^{-1} H'_1$.

An alternative sufficient condition for Assumption 12 is as follows:

ASSUMPTION 12^{2*}: (a) $\hat{V}_{T\pi} = H_1 \hat{\mathscr{J}}_{*T\pi}^{-1} H'_1$.
(b) $\hat{\mathscr{J}}_{*T\pi} = \mathscr{J}_{*T\pi} + o_{p\pi}(1)$.

Under Assumption 12^{2*}, $V_\pi = H_1 \mathscr{J}_{*\pi}^{-1} H'_1$.

The choice of $\hat{V}_{T\pi}$ given in Assumption 12* is often preferable to that in Assumption 12^{2*}, because the asymptotic distribution of the Wald statistic under

Assumption 12* involves fewer nuisance parameters than under Assumption 12^{2^*} when Assumption 11(a) holds, but Assumption 11(b) does not; see the Comment to Theorem 5.

6.2. *The Asymptotic Null Distribution of the Wald Statistic*

The asymptotic null distribution of W_T is given in the following theorem.

THEOREM 6: *Suppose* $\theta_0 \in \Theta_0$ *and Assumptions* 2–9 *and* 12 *hold. Then*:
(a) $W_T(\pi) = \hat{\lambda}'_{\beta_1 T\pi} \hat{V}_{T\pi}^{-1} \hat{\lambda}_{\beta_1 T\pi} + o_{p\pi}(1)$,
(b) $W_T = \sup_{\pi \in \Pi} \hat{\lambda}'_{\beta_1 T\pi} \hat{V}_{T\pi}^{-1} \hat{\lambda}_{\beta_1 T\pi} + o_p(1)$,
(c) $W_T(\bullet) \Rightarrow \hat{\lambda}'_{\beta_1 \bullet} V_\bullet^{-1} \hat{\lambda}_{\beta_1 \bullet}$, *and*
(d) $W_T \to_d AD_W := \sup_{\pi \in \Pi} \hat{\lambda}'_{\beta_1 \pi} V_\pi^{-1} \hat{\lambda}_{\beta_1 \pi} \geq 0$, *where* $\hat{\lambda}_{\beta\pi} = (\hat{\lambda}'_{\beta_1 \pi}, \hat{\lambda}'_{\beta_2 \pi})'$, $\hat{\lambda}_{\beta\pi} \in \mathrm{cl}(\Lambda_\beta)$, *and* $\hat{\lambda}_{\beta\pi}$ *satisfies* $q_\beta(\hat{\lambda}_{\beta\pi}, \pi) = \inf_{\lambda_\beta \in \Lambda_\beta} q_\beta(\lambda_\beta, \pi)$.

COMMENTS: 1. In comparison to Theorem 4 for the QLR statistic, Theorem 6 requires that Λ be convex (Assumption 6), but does not require Assumptions 4_0 or 10. Assumptions 4_0 and 6 hold in most applications. Assumption 10(a), however, does not always hold.

2. The statistics QLR_T and W_T have the same asymptotic null distribution (i.e., $AD_{QLR} = AD_W$) if $\hat{V}_{T\pi} = H_1 \mathscr{I}_{*T\pi}^{-1} H_1' + o_{p\pi}(1)$ and either (i) $p_2 = 0$ or (ii) $H \mathscr{I}_{*\pi}^{-1} H'$ is block diagonal with $p_1 \times p_1$ and $p_2 \times p_2$ blocks, with lower block that does not depend on π, and $Z_{\beta_2 \pi}$ does not depend on π. The condition on $\hat{V}_{T\pi}$ holds under Assumption 12^{2^*}. It also holds under Assumptions 11 and 12* when $c_\pi = 1$. The statistics $RQLR_T$ and W_T have the same asymptotic null distribution (i.e., $AD_{RQLR} = AD_W$) if $\hat{V}_{T\pi} = H_1 \mathscr{I}_{*T\pi}^{-1} \mathscr{I}_{*T\pi} \mathscr{I}_{*T\pi}^{-1} H_1' + o_{p\pi}(1)$ and either condition (i) or (ii) above holds. The condition on $\hat{V}_{T\pi}$ holds under Assumption 12*.

One might think of defining a Wald statistic that has the same asymptotic distribution as the QLR_T statistic whether or not condition (i) or (ii) holds by basing it on the difference between quadratic forms in $\hat{\beta}_\pi$ and $\hat{\beta}_{0\pi}$, where $\hat{\theta}_{0\pi} = (\hat{\beta}'_{0\pi}, \hat{\delta}'_{0\pi}, \hat{\psi}'_{0\pi})'$ and $\hat{\beta}_{0\pi} = (\beta'_{1*}, \hat{\beta}'_{20\pi})'$. This does not work, however, because $\hat{\beta}_{2\pi}$ and $\hat{\beta}_{20\pi}$ need to be centered at β_{20} (where $\theta_0 = (\beta'_{1*}, \beta'_{20}, \delta'_0, \psi'_0)'$ under H_0) and β_{20} is unknown.

3. If Assumption 11 holds, Λ_β is convex, and $\hat{V}_{T\pi} = H_1 \mathscr{I}_{*T\pi}^{-1} \mathscr{I}_{*T\pi} \mathscr{I}_{*T\pi}^{-1} H_1' + o_{p\pi}(1)$, then the distribution of $\hat{\lambda}'_{\beta_1 \pi} V_\pi^{-1} \hat{\lambda}_{\beta_1 \pi}$ for fixed π is that of a mixture of chi-square random variables; see Shapiro (1985, Thm. 3.1 and Sec. 4).

6.3. *GARCH Example* (*Continued*)

We choose the weight matrix of the Wald statistic to eliminate any nuisance parameters in the asymptotic null distribution of the statistic and to be as simple as possible. In particular, we take $\hat{V}_{T\pi} := H_1 \mathscr{I}_{*\pi}^{-1} \mathscr{I}_{*\pi,\pi} \mathscr{I}_{*\pi}^{-1} H_1' = 1 - \pi^2$. This choice satisfies Assumption 12*. The requirement of (2.2) that $\sup\{\pi \in \Pi\} < 1$

ensures that Assumption 12*(e) holds. With this choice of weight matrix, we have

$$(6.3) \qquad W_T := T \sup_{\pi \in \Pi} \hat{\beta}_{1\pi}^2 / (1 - \pi^2).$$

By Theorem 6(d), under the null hypothesis,

$$(6.4) \qquad W_T \to^d \sup_{\pi \in \Pi} \hat{\lambda}_{\beta_1 \pi} / (1 - \pi^2) = \sup_{\pi \in \Pi} \max{}^2\{Z_{\beta_1 \pi}, 0\} / (1 - \pi^2)$$
$$\simeq \sup_{\pi \in \Pi} \max{}^2\{v_\pi, 0\},$$

where v_π is defined in (5.15). Note that the asymptotic null distribution of the Wald statistic is nuisance parameter free. It is the same as that of the *RQLR* statistic. Critical values are given in Section 5.5.

6.4. *Random Coefficient Example* (*Continued*)

We take the weight matrix to be $\hat{V}_{T\pi} := H_1 \hat{\mathscr{J}}_{*T\pi}^{-1} \hat{\mathscr{I}}_{*T\pi} \hat{\mathscr{J}}_{*T\pi}^{-1} H_1'$, where $\hat{\mathscr{J}}_{*T\pi}$ and $\hat{\mathscr{I}}_{*T\pi}$ are as defined in (5.19) and (5.20) using the unrestricted estimator $\hat{\theta}_\pi$ in place of $\hat{\theta}_0$. With this choice, Assumption 12*(a)–(d) holds. Assumption 12*(e) holds because $\mathscr{I}_{*\pi,\pi}$, defined in (4.15), is positive definite by (4.14) and continuous in π and Π is compact. By Theorem 6(d), W_T has the same asymptotic null distribution as $RQLR_T$; see (5.21).

7. THE SCORE TEST

7.1. *The Directed Score and Score Statistic*

In this section, we introduce a score test. The score test statistic is defined to be a quadratic form in a vector of "directed scores," denoted ds_π. The directed score vector is the part of the score of the estimator objective function that is relevant to the null hypothesis, evaluated at the restricted estimator $\hat{\theta}_{0\pi}$, and directed to lie in the parameter space.

The weight matrix $\hat{V}_{T\pi}$ for the score test can be taken as in Assumptions 12, 12*, or 12^{2^*}. Usually, one evaluates the weight matrix at the restricted estimator $\hat{\theta}_{0\pi}$ for the score statistic and at the unrestricted estimator $\hat{\theta}_\pi$ for the Wald test (because then the score statistic does not require that one calculate $\hat{\theta}_\pi$ and the Wald statistic does not require that one calculate $\hat{\theta}_{0\pi}$). Assumptions 12, 12*, and 12^{2^*}, however, do not distinguish between these two cases. Either is permitted. Thus, we employ these assumptions as they are stated in Section 6.

We start by introducing the score function for the parameter $\theta_* = (\beta', \delta')'$. We suppose that there exists a random function $D_* \ell_T(\theta, \pi) \in R^{p+q}$, which we call the *score function*, such that

$$(7.1) \qquad D_* \ell_T(\theta, \pi) := D_* \ell_T(\theta_0, \pi) + D_*^2 \ell_T(\theta_0, \pi)(\theta - \theta_0) + R_T^*(\theta, \pi),$$

where $D_*\ell_T(\theta_0,\pi)$ and $D_*^2\ell_T(\theta_0,\pi)$ equal the first $p+q$ rows of $D\ell_T(\theta_0,\pi)$ and $D^2\ell_T(\theta_0,\pi)$ of (3.3), respectively, and $R_T^*(\theta,\pi)$ is an R^{p+q}-valued random remainder term that satisfies Assumption 13 below. If $\ell_T(\theta,\pi)$ has pointwise partial derivatives with respect to (wrt) θ_*, then $D_*\ell_T(\theta,\pi)$ equals the vector of partial derivatives of $\ell_T(\theta,\pi)$ wrt θ_*. As with Assumption 2, however, we do not require that $\ell_T(\theta,\pi)$ has pointwise partial derivatives. Our results allow $\ell_T(\theta,\pi)$ and $D_*\ell_T(\theta,\pi)$ to have kinks and discontinuities as functions of θ. In this respect, our results are novel even in the classical special case where the estimator objective function does not depend on π and the parameter space Θ contains a neighborhood of θ_0. We are not aware of any papers that consider score tests or LM tests with nondifferentiable estimator objective functions.

The directed score depends on an estimator $\hat{\mathscr{J}}_{*T\pi}$ of $\mathscr{J}_{*T\pi}$ (for $\mathscr{J}_{*T\pi}$ as in (6.2)).

ASSUMPTION 13: (a) *For all* $0<\gamma<\infty$, $\sup_{\theta\in\Theta_0:\|B_T(\theta-\theta_0)\|\le\gamma}\|B_{*T}^{-1\prime}R_T^*(\theta,\pi)\| = o_{p\pi}(1)$.
(b) $\hat{\mathscr{J}}_{*T\pi}=\mathscr{J}_{*T\pi}+o_{p\pi}(1)$.

LEMMA 3: *Assumption* 2^{2^*} *implies Assumption* 13(*a*).

The Appendix provides an alternative sufficient condition for Assumption 13(a) that utilizes stochastic differentiability rather than pointwise smoothness of $D_*\ell_T(\theta,\pi)$.

The *directed score* ds_π is defined by $ds_\pi\in B_{\beta_1T}(\mathscr{B}_1-\beta_{1*})$ and

$$\hat{q}_{\beta_1T}(ds_\pi,\pi)=\inf_{\lambda_{\beta_1}\in B_{\beta_1T}(\mathscr{B}_1-\beta_{1*})}\hat{q}_{\beta_1T}(\lambda_{\beta_1},\pi)+o_{p\pi}(1),\qquad\text{where}\tag{7.2}$$

$$\hat{q}_{\beta_1T}(\lambda_{\beta_1},\pi):=\left(\lambda_{\beta_1}-H_1\hat{\mathscr{J}}_{*T\pi}^{-1}B_{*T}^{-1\prime}D_*\ell_T\left(\hat{\theta}_{0\pi},\pi\right)\right)'\left(H_1\hat{\mathscr{J}}_{*T\pi}^{-1}H_1'\right)^{-1}\times\left(\lambda_{\beta_1}-H_1\hat{\mathscr{J}}_{*T\pi}^{-1}B_{*T}^{-1\prime}D_*\ell_T\left(\hat{\theta}_{0\pi},\pi\right)\right)\qquad\text{and}$$

$$H_1=[I_{p_1}:\mathbf{0}]\in R^{p_1\times(p+q)}.$$

The parameter space $\mathscr{B}_1$ that is used to define the directed score is defined in Assumption 9(c). It is the parameter space for the subvector β_1 of θ. Thus, if Θ is a product set $\mathscr{B}_1\times\mathscr{B}_2\times\Delta\times\Psi$, then $\mathscr{B}_1$ is the set used to define the directed score.

If $\mathscr{B}_1=R^{p_1}$, then the directed score is undirected and equals $H_1\hat{\mathscr{J}}_{*T\pi}^{-1}B_{*T}^{-1\prime}\cdot D_*\ell_T(\hat{\theta}_{0\pi},\pi)$. The latter is just the part of the normalized score function that relates to β_1—the parameter of interest—evaluated at the restricted estimator of θ, $\hat{\theta}_{0\pi}$. When $\mathscr{B}_1\neq R^{p_1}$, then the directed score is defined so that it only takes values that $B_{\beta_1T}(\hat{\beta}_{1\pi}-\beta_{1*})$ can take. That is, it only takes values in $B_{\beta_1T}(\mathscr{B}_1-\beta_{1*})$.

If $B_{\beta_1T}(\mathscr{B}_1-\beta_{1*})=R^+$, then ds_π is given by (3.10) with $\hat{\lambda}_{\beta\pi}$ and $Z_{\beta\pi}$ replaced by ds_π and $H_1\hat{\mathscr{J}}_{*T\pi}^{-1}B_{*T}^{-1\prime}D_*\ell_T(\hat{\theta}_{0\pi},\pi)$. If $B_{\beta_1T}(\mathscr{B}_1-\beta_{1*})$ is a cone defined by linear inequality and/or equality constraints, then a closed form expression for ds_π is given by Theorem 5 or (6.6) of E1 with $\hat{\lambda}_\beta, Z_\beta, H\mathscr{J}_*^{-1}H'$, and Λ_β replaced by ds_π, $H_1\hat{\mathscr{J}}_{*T\pi}^{-1}B_{*T}^{-1'}D_*\ell_T(\hat{\theta}_{0\pi},\pi)$, $H_1\hat{\mathscr{J}}_{*T\pi}^{-1}H_1'$, and $B_{\beta_1 T}(\mathscr{B}_1-\beta_{1*})$ respectively.

The *score test statistic*, S_T, is defined by

$$(7.3)\qquad S_T := \sup_{\pi\in\Pi} S_T(\pi)+o_p(1), \qquad \text{where} \qquad S_T(\pi) := ds_\pi'\hat{V}_{T\pi}^{-1}\,ds_\pi$$

and $\hat{V}_{T\pi}$ satisfies Assumption 12.

7.2. *The Asymptotic Null Distribution of the Score Statistic*

We define

$$(7.4)\qquad q_{\beta_1 T}(\lambda_{\beta_1},\pi) := (\lambda_{\beta_1}-Z_{\beta_1 T\pi})'(H_1\mathscr{J}_{*T\pi}^{-1}H_1')^{-1}(\lambda_{\beta_1}-Z_{\beta_1 T\pi}) \qquad \text{and}$$

$$Z_{T\pi} := (Z_{\beta_1 T\pi}', Z_{\beta_2 T\pi}', Z_{\delta T\pi}', Z_{\psi T\pi}')'$$

for $Z_{T\pi}$ as defined in (3.3).

The asymptotic properties of ds_π are given in the following lemma.

LEMMA 4: *Suppose* $\theta_0\in\Theta_0$ *and Assumptions* 2, 3, 4_0, 5–9, *and* 13 *hold*. *Then*,

(a) $H_1\hat{\mathscr{J}}_{*T\pi}^{-1}B_{*T}^{-1'}D_*\ell_T(\hat{\theta}_{0\pi},\pi) = H_1\mathscr{J}_{*T\pi}^{-1}B_{*T}^{-1'}D_*\ell_T(\theta_0,\pi)+o_{p\pi}(1) = Z_{\beta_1 T\pi}+o_{p\pi}(1)$,

(b) $ds_\pi = O_{p\pi}(1)$,

(c) $\hat{q}_{\beta_1 T}(ds_\pi,\pi) = q_{\beta_1 T}(ds_\pi,\pi)+o_{p\pi}(1)$,

(d) $q_{\beta_1 T}(ds_\pi,\pi) = \inf_{\lambda_{\beta_1}\in B_{\beta_1 T}(\mathscr{B}_1-\beta_{1*})} q_{\beta_1 T}(\lambda_{\beta_1},\pi)+o_{p\pi}(1)$,

(e) $\inf_{\lambda_{\beta_1}\in B_{\beta_1 T}(\mathscr{B}_1-\beta_{1*})} q_{\beta_1 T}(\lambda_{\beta_1},\pi) = \inf_{\lambda_{\beta_1}\in\Lambda_{\beta_1}} q_{\beta_1 T}(\lambda_{\beta_1},\pi)+o_{p\pi}(1)$,

(f) $q_{\beta_1 T}(\hat{\lambda}_{\beta_1 T\pi},\pi) = \inf_{\lambda_{\beta_1}\in\Lambda_{\beta_1}} q_{\beta_1 T}(\lambda_{\beta_1},\pi)+o_{p\pi}(1)$ *for* $\hat{\lambda}_{\beta_1 T\pi}$ *defined in* (6.2) *and*

(g) $ds_\pi = \hat{\lambda}_{\beta_1 T\pi}+o_{p\pi}(1)$.

COMMENT: Lemma 4(g) and Theorem 6(a) combine to show that the Wald and score test statistics are asymptotically equivalent (i.e., $W_T = S_T+o_{p\pi}(1)$ for $\theta_0\in\Theta_0$) when they are constructed using asymptotically equivalent weight matrices. Furthermore, by Theorem 4(a) and the proof of Theorem 4 of E1 adjusted appropriately, we obtain

$$(7.5)\qquad QLR_T = \sup_{\pi\in\Pi}\hat{\lambda}_{\beta T\pi}'(H\mathscr{J}_{*T\pi}^{-1}H')^{-1}\hat{\lambda}_{\beta T\pi} - \sup_{\pi\in\Pi}\hat{\lambda}_{\beta 0T\pi}'(H\mathscr{J}_{*T\pi}^{-1}H')^{-1}\hat{\lambda}_{\beta 0T\pi}+o_p(1),$$

where $\hat{\lambda}_{0T\pi} = (\hat{\lambda}_{\beta 0T\pi}', \hat{\lambda}_{\delta 0T\pi}', \hat{\lambda}_{\psi 0T\pi}')'$ is defined to satisfy $\hat{\lambda}_{0T\pi}\in \mathrm{cl}(\Lambda_0)$ and $q_T(\hat{\lambda}_{0T\pi},\pi) = \inf_{\lambda\in\Lambda_0} q_T(\lambda,\pi)$, under Assumptions 2–5, 4_0, and 7–10 when $\theta_0\in\Theta_0$. Thus, the QLR, Wald, and score test statistics are asymptotically equivalent for $\theta_0 \in \Theta_0$ whenever the weight matrix $V_{T\pi}^{-1}$ of the latter two statistics $V_{T\pi} = H_1 \mathscr{J}_{*T\pi}^{-1} H_1' + o_{p\pi}(1)$ and either (i) $p_2 = 0$ or (ii) $H \mathscr{J}_{*T\pi}^{-1} H'$ is block diagonal with $p_1 \times p_1$ and $p_2 \times p_2$ blocks, with lower block that does not depend on π, and $Z_{\beta_2 T\pi}$ does not depend on π (because if (i) holds, then $\hat{\lambda}'_{\beta 0 T\pi} = 0$, and if (ii) holds, then $\hat{\lambda}_{\beta_2 0 T\pi} = \hat{\lambda}_{\beta_2 T\pi}$, $\hat{\lambda}_{\beta_2 T\pi}$ does not depend on π, and $\hat{\lambda}_{\beta_1 0 T\pi} = 0$).

The asymptotic null distribution of the score statistic S_T is as follows.

THEOREM 7: *Suppose* $\theta_0 \in \Theta_0$ *and Assumptions* 2, 3, 4_0, 5–9, 12, *and* 13 *hold*. *Then*:

(a) $S_T(\pi) = \hat{\lambda}'_{\beta_1 T\pi} \hat{V}_{T\pi}^{-1} \hat{\lambda}_{\beta_1 T\pi} + o_{p\pi}(1)$,

(b) $S_T = \sup_{\pi \in \Pi} \hat{\lambda}'_{\beta_1 T\pi} \hat{V}_{T\pi}^{-1} \hat{\lambda}_{\beta_1 T\pi} + o_p(1)$,

(c) $S_T(\bullet) \Rightarrow \hat{\lambda}'_{\beta_1 \bullet} V_{\bullet}^{-1} \hat{\lambda}_{\beta_1 \bullet}$, *and*

(d) $S_T \to_d AD_S := \sup_{\pi \in \Pi} \hat{\lambda}'_{\beta_1 \pi} V_\pi^{-1} \hat{\lambda}_{\beta_1 \pi}$, *where* $\hat{\lambda}_{\beta_1 \pi}$ *is as in Theorem* 6.

COMMENTS: 1. The score test does not require one to compute the unrestricted estimator $\hat{\theta}_\pi$. This has computational advantages in some cases over the QLR and Wald tests, which require computation of $\hat{\theta}_\pi$.

2. The choice of weight matrix to satisfy Assumption 12* or 12^{2^*} has the same effect on the asymptotic distribution of the score statistic as it does on the asymptotic distribution of the Wald statistic, as discussed above.

7.3. *GARCH Example* (*Continued*)

In this example, the score function $D_* \ell_T(\theta, \pi)$ and $\hat{\mathscr{J}}_{*T\pi}$ are

$$D_* \ell_T(\theta, \pi) := \sum_{t=1}^{T} \left(\frac{1}{2} (e_t^2(\theta) - h_t^*(\theta, \pi)) \frac{\partial}{\partial \theta_*} h_t^*(\theta, \pi) / (h_t^*(\theta, \pi))^2 \right) \tag{7.6}$$

and

$$\hat{\mathscr{J}}_{*T\pi} := \frac{1}{2} \begin{pmatrix} 2\hat{c}_0/(1-\pi^2) + 1/(1-\pi)^2 & 1/\big((1-\pi)\hat{\delta}_0\big) \\ 1/\big((1-\pi)\hat{\delta}_0\big) & 1/\hat{\delta}_0^2 \end{pmatrix},$$

where

$$\hat{c}_0 := \hat{\tau}_0 / \big(2\hat{\delta}_0^2\big) \qquad \text{and} \qquad \hat{\tau}_0 := \frac{1}{T} \sum_{t=1}^{T} \big(\hat{e}_t^2 - \hat{\delta}_0\big)^2.$$

Assumption 13(a) holds by Lemma 3(a) because Assumption 2^{2^*} is verified in the Appendix. Assumption 13(b) holds using the definitions of $\mathscr{J}_{*T\pi} (= \mathscr{J}_{*\pi})$ and $\hat{\mathscr{J}}_{*T\pi}$ in (4.3), (5.13), and (7.6), the moment conditions in (2.3), the law of large numbers for stationary and ergodic processes, and some simple manipulations.

Some calculations give

$$(7.7) \qquad H\hat{\mathscr{J}}_{*T\pi}^{-1}B_{*T}^{-1'}D_{*}\ell_T\left(\hat{\theta}_{0\pi},\pi\right) = \frac{(1-\pi^2)}{\hat{\tau}_0 T^{1/2}}\sum_{t=1}^{T}\left(\hat{e}_t^2-\hat{\delta}_0\right)\sum_{k=0}^{t-2}\pi^k\hat{e}_{t-k-1}^2.$$

(We assume $h_1^*(\theta,\pi)$ does not depend on β_1 when defining $D_*\ell_T(\theta,\pi)$. This implies that the term $\pi^{t-1}\partial h_1^*(\theta,\pi)/\partial\beta_1$, which otherwise would appear in the formula for $\partial h_t^*(\theta,\pi)/\partial\beta_1$, is zero.)

The set $B_{\beta_1 T}(\mathscr{B}_1-\beta_{1*})$ equals R^+ in the present case. In consequence, a closed form solution for ds_π can be obtained quite easily:

$$(7.8) \qquad ds_\pi = \max\left\{\frac{(1-\pi^2)}{\hat{\tau}_0 T^{1/2}}\sum_{t=1}^{T}\left(\hat{e}_t^2-\hat{\delta}_0\right)\sum_{k=0}^{t-2}\pi^k\hat{e}_{t-k-1}^2, 0\right\}.$$

We take $\hat{V}_{T\pi} = H\mathscr{J}_{*\pi}^{-1}\mathscr{I}_{*\pi,\pi}\mathscr{J}_{*\pi}^{-1}H' = 1-\pi^2$. This choice of weight matrix satisfies Assumption 12*. With this weight matrix the score test statistic is

$$(7.9) \qquad S_T := \sup_{\pi\in\Pi}(1-\pi^2)\max{}^2\left\{\frac{1}{\hat{\tau}_0 T^{1/2}}\sum_{t=1}^{T}\left(\hat{e}_t^2-\hat{\delta}_0\right)\sum_{k=0}^{t-2}\pi^k\hat{e}_{t-k-1}^2, 0\right\}.$$

The asymptotic null distribution of the score statistic is given by Theorem 7(d). It is the same as that of the Wald statistic; see (6.4).

7.4. *Random Coefficient Example* (*Continued*)

The function $D_*\ell_T(\theta,\pi)$ in this example is

$$(7.10) \qquad D_*\ell_T(\theta,\pi) := \sum_{t=1}^{T}\frac{\text{res}_t^2(\theta)-\text{var}_t(\theta,\pi)}{2\text{var}_t^2(\theta,\pi)}\tilde{W}_t(\theta,\pi),$$

where $\tilde{W}_t(\theta,\pi)$ is defined in (5.19). The matrix $\hat{\mathscr{J}}_{*T\pi}$ of Assumption 13(b) is defined in (5.19). Assumption 13(b) holds by Assumption 1_0, a uniform law of large numbers, and the continuity of $E\tilde{W}_t(\theta,\pi)\tilde{W}_t(\theta,\pi)'/\text{var}_t^2(\theta,\pi)$ in θ. Assumption 13(a) holds by Lemma 3 because Assumption 2^{2^*} is verified in the Appendix.

We take $\hat{V}_{T\pi} := H\hat{\mathscr{J}}_{*T\pi}^{-1}\hat{\mathscr{I}}_{*T\pi}(\hat{\theta}_0)\mathscr{J}_{*T\pi}^{-1}H'$, where $\hat{\mathscr{I}}_{*T\pi}$ is defined in (5.20). This choice of weight matrix satisfies Assumption 12*. (Note that the weight matrix is actually a scalar in this case because $p=1$.)

We have $B_{\beta_1 T}(\mathscr{B}_1-\beta_{1*}) = R^+$ and a closed form expression for ds_π is

$$(7.11) \qquad ds_\pi := \max\left\{H\hat{\mathscr{J}}_{*T\pi}^{-1}T^{-1/2}D_*\ell_T\left(\hat{\theta}_0,\pi\right), 0\right\}.$$

The score test statistic is

$$(7.12) \qquad S_T := \sup_{\pi\in\Pi}\max{}^2\left\{\hat{V}_{T\pi}^{-1/2}T^{-1/2}\sum_{t=1}^{T}\frac{\text{res}_t^2\left(\hat{\theta}_0\right)-\text{var}_t\left(\hat{\theta}_0\right)}{2\text{var}_t^2\left(\hat{\theta}_0\right)} \times H\hat{\mathscr{J}}_{*T\pi}^{-1}\tilde{W}_t\left(\hat{\theta}_0,\pi\right), 0\right\}.$$

The asymptotic null distribution of the score statistic is given by Theorem 7(d). It is the same as that of the RQLR and Wald statistics; see (5.21).

8. LOCAL POWER

In this section, we consider the asymptotic distributions of the QLR, RQLR, Wald, and score tests under sequences of local alternatives to the null parameter value θ_0. We consider sequences of pseudo-true values of the form $\theta_{T0} = \theta_0 + B_T^{-1}\eta_T$, where $\{\eta_T : T \geq 1\}$ is a sequence of constant s-vectors that satisfies $\eta_T \to \eta$ for some η. When the data are generated by such a sequence, we say that the data are generated "under θ_{T0}."

One can use the results of Sections 3 and 6–7 to determine the desired asymptotic results under the local alternatives θ_{T0}. One verifies Assumptions 1, 1_0, 2, 4–6, 4_0, 7(b), 8, 9, and 13 (and all the superscripted versions of these assumptions) under θ_{T0} with exactly the same quantities appearing in these assumptions as under θ_0. For example, one verifies Assumption 1 under θ_{T0} with the probability limit being θ_0 just as under θ_0 and one verifies Assumption 2 under θ_{T0} with the components of the quadratic approximation being the same as under θ_0.

Next, one verifies Assumption 3 or 3* with the same processes $(B_T^{-1'}D\ell_T(\theta_0, \bullet), \mathscr{J}_{T\bullet})$ under θ_{T0} as under θ_0, but with a different limit process $(G_\bullet, \mathscr{J}_\bullet)$ under θ_{T0} than under θ_0. For example, if $G_\bullet$ is a mean zero Gaussian process under θ_0, then $G_\bullet$ typically is a Gaussian process with the same covariance function but a nonzero mean under θ_{T0}. The asymptotic distributions of the test statistics differ under θ_{T0} than under θ_0 because the limit process $(G_\bullet, \mathscr{J}_\bullet)$ differs. In consequence, the tests typically have nontrivial asymptotic power against local alternatives. Note that for cases without stochastic trends $\mathscr{J}_\bullet$ of Assumption 3 typically is nonrandom and is the same under θ_{T0} as under θ_0.

As with Assumption 3, one verifies Assumptions 7(a) and 10–12 under θ_{T0} with limit processes $\mathscr{J}_\bullet$, $G_{\delta\bullet}$, $G_{\psi\bullet}, \ldots, c_\bullet$, and $V_\bullet$ that may be different under θ_{T0} than under θ_0. In many cases, however, $\mathscr{J}_\bullet$, $c_\bullet$, and $V_\bullet$ are nonrandom and are the same under θ_{T0} as under θ_0.

Assumptions 5, 6, 7(b), 8, 9, and 10(a) do not depend on the distribution of the data and, hence, if they have been verified for results under θ_0 they also hold for results under the local alternatives θ_{T0}.

In many cases, the local alternatives θ_{T0} are contiguous to the null θ_0. In such cases, if Assumptions 1, 1_0, 2, and 13 (and any of the superscripted versions of these assumptions) hold under θ_0, then they automatically hold under θ_{T0}. Furthermore, if Assumptions 7(a), 11(b), 11(c), and 12 hold under θ_0 with nonrandom limit processes $\mathscr{J}_\bullet$, $\mathscr{I}_{*\bullet}$, $c_\bullet$, and $V_\bullet$, as is typically the case when no stochastic trends are present, then they automatically hold under θ_{T0}. If Assumptions 4 and 4_0 are verified using Lemma 1 under θ_0, then they also hold via Lemma 1 under θ_{T0} provided Assumption 3 holds under θ_{T0}. Thus, with contiguous local alternatives, the main task is to verify Assumption 3 under θ_{T0}. This can be done using the same methods as when verifying it under θ_0.

For brevity, we do not provide local power results for the two examples.

Cowles Foundation for Research in Economics, Yale University, 30 Hillhouse Ave., New Haven, CT. 06520-8281, U.S.A.; donald.andrews@yale.edu

Manuscript received November, 1998; final revision received March, 2000.

APPENDIX OF PROOFS

A. Sufficient Conditions for Assumptions 2* *and* 4

The form of Assumption 2* is such that one can replace the objective function $\ell_T(\theta,\pi)$ by a more tractable function, say $\mathscr{L}_T(\theta,\pi)$, that is a close approximation to $\ell_T(\theta,\pi)$. For example, in the GARCH(1,1) Example, $\ell_T(\theta,\pi)$ is a sum of quasi-log likelihood contributions that depends on initial conditions and, hence, is not stationary and ergodic. We can define a more tractable function $\mathscr{L}_T(\theta,\pi)$ to be the stationary and ergodic analogue of $\ell_T(\theta,\pi)$ that replaces the initial conditions by terms that depend on the infinite history of the process. Now, suppose

$$(9.1)\qquad \sup_{\theta\in\Theta:\|\theta-\theta_0\|\le\gamma_T,\,\pi\in\Pi} |\ell_T(\theta,\pi)-\ell_T(\theta_0,\pi)-\mathscr{L}_T(\theta,\pi)+\mathscr{L}_T(\theta_0,\pi)| = o_p(1)$$

for all $\gamma_T\to 0$. Also, suppose $\mathscr{L}_T(\theta,\pi)$ has an expansion of the form

$$(9.2)\qquad \mathscr{L}_T(\theta,\pi)=\mathscr{L}_T(\theta_0,\pi)+D\mathscr{L}_T(\theta_0,\pi)'(\theta-\theta_0) + \frac{1}{2}(\theta-\theta_0)'D^2\mathscr{L}_T(\theta_0,\pi)(\theta-\theta_0)+R_T^*(\theta,\pi),$$

where $R_T^*(\theta,\pi)$ satisfies Assumption 2* with $R_T(\theta,\pi)$ replaced by $R_T^*(\theta,\pi)$. Then, $\ell_T(\theta,\pi)$ satisfies (3.3) with

$$(9.3)\qquad D\ell_T(\theta_0,\pi)=D\mathscr{L}_T(\theta_0,\pi),\qquad D^2\ell_T(\theta_0,\pi)=D^2\mathscr{L}_T(\theta_0,\pi),\qquad \text{and}$$
$$R_T(\theta,\pi)=R_T^*(\theta,\pi)+(\ell_T(\theta,\pi)-\ell_T(\theta_0,\pi)-\mathscr{L}_T(\theta,\pi)+\mathscr{L}_T(\theta_0,\pi)).$$

Assumption 2* holds for $\ell_T(\theta,\pi)$ in this case by (9.1) and (9.2).

Next, we give a sufficient condition for Assumption 2* that does not rely on smoothness of $\ell_T(\theta,\pi)$ in θ. We say that a sequence of random functions $\{g_T(\theta,\pi):T\ge 1\}$ is *stochastically differentiable* at θ_0 for $\Theta\subset R^s$ uniformly over Π with random derivative s-vector $Dg_T(\theta_0,\pi)$ if

$$(9.4)\qquad g_T(\theta,\pi)=g_T(\theta_0,\pi)+Dg_T(\theta_0,\pi)'(\theta-\theta_0)+r_T(\theta,\pi)\qquad\text{and}$$
$$\sup_{\theta\in\Theta:\|\theta-\theta_0\|\le\gamma_T} T|r_T(\theta,\pi)|/(1+\|T^{1/2}(\theta-\theta_0)\|)=o_{p\pi}(1)$$

for all $\gamma_T\to 0$. We apply the concept of stochastic differentiability to random functions $g_T(\theta,\pi)$ that are $O_{p\pi}(1)$, such as sample averages, and for which $T^{1/2}Dg_T(\theta_0,\pi)=O_{p\pi}(1)$.

Assumption 2^{3^*}: (a) $B_T=T^{1/2}I_s$.

(b) *For some nonrandom function* $\ell(\theta,\pi)$, $T^{-1}\ell_T(\theta,\pi)\to^p \ell(\theta,\pi)$ $\forall\theta\in\Theta\cap S(\theta_0,\varepsilon)$ *for some* $\varepsilon>0$, $\forall\pi\in\Pi$.

(c) *For each* $\pi\in\Pi$, *the domain of* $\ell(\theta,\pi)$ *as a function of* θ *includes a set* Θ^+ *that satisfies conditions* (i) *and* (ii) *of Assumption* $2^{2^*}K(a)$.

(d) $\ell(\theta,\pi)$ *has continuous* ℓ/r *partial derivatives with respect to* θ *of order two on* Θ^+ $\forall\pi\in\Pi$, *with* ℓ/r *partial derivatives* $\partial\ell(\theta,\pi)/\partial\theta$ *and* $\partial^2\ell(\theta,\pi)/\partial\theta\,\partial\theta'$ *of orders one and two, respectively, that satisfy* $\partial\ell(\theta_0,\pi)/\partial\theta=\mathbf{0}$ $\forall\pi\in\Pi$ *and* $\partial^2\ell(\theta,\pi)/\partial\theta\,\partial\theta'$ *is continuous at* θ_0 *uniformly over* $\pi\in\Pi$.

(e) $\{T^{-1}\ell_T(\theta,\pi)-\ell(\theta,\pi):T\ge 1\}$ *is stochastically differentiable at* θ_0 *for* Θ *uniformly over* Π *with random derivative vector* $T^{-1}D\ell_T(\theta_0,\pi)$.

The proof of the sufficiency of Assumption 2^{3^*} for Assumption 2^* is analogous to that given in E2 for the case where no parameter π appears.

The empirical process results referred to in E2 can be used to verify the stochastic differentiability assumption, Assumption 2^{3^*}(e).

B. Sufficient Conditions for Assumption 13

We now introduce a sufficient condition for Assumption 13 that uses stochastic differentiability and allows $D_* \ell_T(\theta, \pi)$ to have kinks and discontinuities. This condition is similar to Assumption 2^{3^*}.

ASSUMPTION 13*: (a) $B_T = T^{1/2} I_s$.

(b) *There exist random R^{p+q}-valued functions $\{D_* \ell_T(\theta, \pi): T \geq 1\}$ and a nonrandom R^{p+q}-valued function $D_* \ell(\theta, \pi)$ that satisfy $T^{-1} D_* \ell_T(\theta, \pi) = D_* \ell(\theta, \pi) + o_{p\pi}(1)$ $\forall \theta \in \Theta \cap S(\theta_0, \varepsilon)$ for some $\varepsilon > 0$.*

(c) *For each $\pi \in \Pi$, the domain of $D_* \ell(\theta, \pi)$ as a function of θ includes a set Θ^+ that satisfies conditions* (i) *and* (ii) *of Assumption 2^{2^*}(a).*

(d) *$D_* \ell(\theta, \pi)$ has continuous ℓ/r partial derivatives with respect to θ of order one on Θ^+ $\forall \pi \in \Pi$ and the ℓ/r partial derivatives, denoted $D_*{}^2 \ell(\theta, \pi) \in R^{(p+q) \times s}$, are continuous at θ_0 uniformly over $\pi \in \Pi$.*

(e) *$\{T^{-1} D_* \ell_T(\theta, \pi) - D_* \ell(\theta, \pi): T \geq 1\}$ is stochastically differentiable at θ_0 for Θ uniformly over Π with random derivative $(p+q) \times s$ matrix $T^{-1} D_*{}^2 \ell_T(\theta_0, \pi) - D_*{}^2 \ell(\theta_0, \pi)$, where $D_*{}^2 \ell_T(\theta_0, \pi)$ equals the first $p+q$ rows of $D^2 \ell_T(\theta_0, \pi)$ of* (3.3).

(f) $\hat{\mathscr{J}}_{*T\pi} = \mathscr{J}_{*T\pi} + o_{p\pi}(1)$.

LEMMA 5: *Assumption* 13* *implies Assumption* 13.

COMMENT: The definition of stochastic differentiability uniformly over Π in (9.4) can be weakened and Lemma 5 still holds. In particular, the T term that appears in the numerator of (9.4) can be replaced by $T^{1/2}$.

C. Proofs of General Results

LEMMA 6: *Under Assumptions* 2–4, $\ell_T(\hat{\theta}_\pi, \pi) = \ell_T(\theta_0, \pi) + \frac{1}{2} Z'_{T\pi} \mathscr{J}_{T\pi} Z_{T\pi} - \frac{1}{2} \inf_{\theta \in \Theta} q_T(B_T(\theta - \theta_0), \pi) + o_{p\pi}(1)$.

LEMMA 7: *Suppose Assumptions* 3 *and* 5 *hold. Then,* $\inf_{\lambda \in B_T(\Theta - \theta_0)} q_T(\lambda, \pi) = \inf_{\lambda \in \Lambda} q_T(\lambda, \pi) + o_{p\pi}(1)$.

PROOF OF LEMMAS 1 AND 6: The proofs are analogous to those of Theorems 1 and 2 of E1, respectively, with $O_p(1)$ and $o_p(1)$ replaced by $O_{p\pi}(1)$ and $o_{p\pi}(1)$ throughout. *Q.E.D.*

PROOF OF LEMMA 7: The proof is analogous to the proof of Lemma 2 of E1 with the subscript π added to $Z_T, Z_{Tb}, \mathscr{J}_T, o_p(1), O_p(1)$, $\text{dist}_T(\bullet, \bullet)$, and C_T and with $q_T(\lambda)$ changed to $q_T(\lambda, \pi)$. No subscript π is added to $\text{dist}(\bullet, \bullet)$, $\|\bullet\|$, or $o(\bullet)$. The subscript π is added to $Z_{\Theta Tb}$ and $Z_{\Lambda Tb}$ except where they appear as $\|Z_{\Theta Tb}\|$ and $\|Z_{\Lambda Tb}\|$, in which case $\|Z_{\Theta Tb}\|$ and $\|Z_{\Lambda Tb}\|$ are defined to equal $\sup_{\pi \in \Pi} \|Z_{\Theta Tb\pi}\|$ and $\sup_{\pi \in \Pi} \|Z_{\Lambda Tb\pi}\|$ respectively. The only exception to the latter is on the left-hand side of the last equation of the proof in which case $\|Z_{\Theta Tb}\|$ denotes $\|Z_{\Theta Tb\pi}\|$. *Q.E.D.*

PROOF OF THEOREM 1: The proof of part (a) is the same as the proof of Theorem 3(a) of E1 with the subscript π added to $\lambda_T^*, \hat{\theta}, \hat{\lambda}_T, Z_T, \|\bullet\|_T, \mathscr{J}_T, o_p(1), \varepsilon_T, \varepsilon_T^*, P_L$, and λ_T^* and with $q_T(\bullet)$ changed to $q_T(\bullet, \pi)$. Given part (a), for part (b) it suffices to show that $\hat{\lambda}_{T\bullet} \Rightarrow \lambda_\bullet$. $\hat{\lambda}_{T\pi}$ is uniquely defined because Λ is a convex cone. We can write $\hat{\lambda}_{T\pi} = h(B_T^{-1'} D\ell_T(\theta_0, \pi), \mathscr{J}_{T\pi})$, where the function h is defined implicitly in (3.4). The function h is uniformly continuous over any set of points $\{(\lambda_c, \mathscr{H}_c): c \in C\}$ for which $0 < \inf\{\lambda_{\min}(\mathscr{H}_c): c \in C\} \le \sup\{\lambda_{\max}(\mathscr{H}_c): c \in C\} < \infty$. Because $0 < \inf\{\lambda_{\min}(\mathscr{J}_\pi): \pi \in \Pi\} \le \sup\{\lambda_{\max}(\mathscr{J}_\pi): \pi \in \Pi\} < \infty$ with $(Z_\bullet, \mathscr{J}_\bullet)$ probability one, the function mapping $(B_T^{-1'} D\ell_T(\theta_0, \bullet), \mathscr{J}_{T\bullet})$ into $\hat{\lambda}_{T\bullet}$ is continuous (with respect to the uniform metric on the space of functions on Π) with $(Z_\bullet, \mathscr{J}_\bullet)$ probability one. Thus, the continuous mapping theorem applies and gives $\hat{\lambda}_{T\bullet} \Rightarrow \lambda_\bullet$.

The convergence in the first result of part (c) holds by Lemmas 6 and 7, Assumption 3, and the continuous mapping theorem. The equality in the first result of part (c) holds by the orthogonality property $\hat{\lambda}'_\pi \mathscr{J}_\pi(\hat{\lambda}_\pi - Z_\pi)$, which does not require Assumption 6; see Perlman (1969, Lemma 4.1), and some algebra. The second result of part (c) holds by the first result and the continuous mapping theorem. *Q.E.D.*

LEMMA 8: *Suppose Assumptions* 3, 7, *and* 8 *hold*. *Then*,
(a) $q_\beta(\hat{\lambda}_{\beta\pi}, \pi) = \inf_{\lambda_\beta \in \Lambda_\beta} q_\beta(\lambda_\beta, \pi)$,
(b) $\hat{\lambda}_{\delta\pi} = \mathscr{J}_{\delta\pi}^{-1} G_{\delta\pi} - \mathscr{J}_{\delta\pi}^{-1} \mathscr{J}_{\delta\beta\pi} \hat{\lambda}_{\beta\pi}$,
(c) $q_\psi(\hat{\lambda}_{\psi\pi}, \pi) = \inf_{\lambda_\psi \in \Lambda_\psi} q_\psi(\lambda_\psi, \pi)$, *and*
(d) $Z'_\pi \mathscr{J}_\pi Z_\pi - \inf_{\lambda \in \Lambda} q(\lambda, \pi) = Z'_{\beta\pi}(H\mathscr{J}_{*\pi}^{-1}H')^{-1} Z_{\beta\pi} - \inf_{\lambda_\beta \in \Lambda_\beta} q_\beta(\lambda_\beta, \pi) + G'_{\delta\pi}\mathscr{J}_{\delta\pi}^{-1} G_{\delta\pi} + Z'_{\psi\pi}\mathscr{J}_{\psi\pi} Z_{\psi\pi} - \inf_{\lambda_\psi \in \Lambda_\psi} q_\psi(\lambda_\psi, \pi) = \hat{\lambda}'_{\beta\pi}(H\mathscr{J}_{*\pi}^{-1}H')^{-1}\hat{\lambda}_{\beta\pi} + G'_{\delta\pi}\mathscr{J}_{\delta\pi}^{-1} G_{\delta\pi} + \hat{\lambda}'_{\psi\pi}\mathscr{J}_{\psi\pi}\hat{\lambda}_{\psi\pi}$.

PROOF OF LEMMA 8: The proof is the same as that of Theorem 4 of E1. *Q.E.D.*

PROOF OF THEOREM 2: Theorem 1 and Lemma 8 combine to establish Theorem 2. *Q.E.D.*

PROOF OF THEOREM 3: To prove part (a), suppose $\phi_{0T} \in B_T(\Theta_0 - \theta_0)/b_T$ $\forall T \ge 1$ and $\|\phi_{0T}\| \to 0$. Because $B_{\beta_1\delta T} = \mathbf{0}$ and $B_{\beta_1\psi T} = \mathbf{0}$, we have $\phi_{\beta_1 0T} = 0$, where $\phi_{0T} = (\phi'_{\beta_1 0T}, \phi'_{\beta_2 0T}, \phi'_{\delta 0T}, \phi'_{\psi 0T})'$. By Assumption 5, there exists $\{\lambda_T \in \Lambda: T \ge 1\}$ such that $\|\phi_{0T} - \lambda_T\| = o(\|\phi_{0T}\|)$. Write $\lambda_T := (\lambda'_{\beta_1 T}, \lambda'_{\beta_2 T}, \lambda'_{\delta T}, \lambda'_{\psi T})'$. Define $\lambda_{0T} := (\mathbf{0}', \lambda'_{\beta_2 T}, \lambda'_{\delta T}, \lambda'_{\psi T})'$. By Assumptions 7–9, $\Lambda = \Lambda_{\beta_1} \times \Lambda_{\beta_2} \times R^q \times \Lambda_\psi$ and, hence, $\lambda_{0T} \in \Lambda_0 := \{\mathbf{0}\} \times \Lambda_{\beta_2} \times R^q \times \Lambda_\psi$. Also, $\|\phi_{0T} - \lambda_{0T}\| \le \|\phi_{0T} - \lambda_T\| = o(\|\phi_{0T}\|)$. Thus, $\mathrm{dist}(\phi_{0T}, \Lambda_0) = o(\|\phi_{0T}\|)$, as desired.

Next, suppose $\lambda_{0T} \in \Lambda_0$ $\forall T \ge 1$ and $\|\lambda_{0T}\| \to 0$. By Assumption 5, there exists $\phi_T \in B_T(\Theta - \theta_0)/b_T$ $\forall T \ge 1$ such that $\|\lambda_{0T} - \phi_T\| = o(\|\lambda_{0T}\|)$. Write $\phi_T := (\phi'_{\beta_1 T}, \phi'_{\beta_2 T}, \phi'_{\delta T}, \phi'_{\psi T})'$. Define $\phi_{0T} := (\mathbf{0}', \phi'_{\beta_2 T}, \phi'_{\delta T}, \phi'_{\psi T})'$. Because Θ is a product set local to θ $\forall \theta \in \Theta_0$, we have $\phi_{0T} \in B_T(\Theta_0 - \theta_0)/b_T$ for T large. Furthermore, $\|\lambda_{0T} - \phi_{0T}\| \le \|\lambda_{0T} - \phi_T\| = o(\|\lambda_{0T}\|)$. Thus, $\mathrm{dist}(\lambda_{0T}, B_T(\Theta_0 - \theta_0)/b_T) = o(\|\lambda_{0T}\|)$ and part (a) is established.

Parts (b) and (c) follow from Lemmas 1 and 6, respectively, with Θ replaced by Θ_0. Part (d) follows from Lemma 7 and Theorem 1(c) with Θ replaced by Θ_0 using the results of parts (a) and (c). Part (e) follows from Lemma 8(d) with Λ, Λ_β, and $\hat{\lambda}_{\beta\pi}$ replaced by Λ_0, $\{\mathbf{0}\} \times \Lambda_{\beta_2}$, and $\hat{\lambda}_{\beta 0\pi}$ respectively. *Q.E.D.*

PROOF OF THEOREM 4: First, by Assumption 10(a), $QLR_T/2 = \sup_{\pi \in \Pi}(\ell_T(\hat{\theta}_\pi, \pi) - \ell_T(\theta_0, \pi)) - \sup_{\pi \in \Pi}(\ell_T(\hat{\theta}_{0\pi}, \pi) - \ell_T(\theta_0, \pi)) + o_p(1)$. Now, part (a) of the Theorem follows from Theorems 3(a) and 3(c) and Lemmas 6 and 7. Part (b) follows from Theorems 1(c) and 3(d). Part (c) follows from Lemma 8(d) and Theorem 3(e). *Q.E.D.*

PROOF OF THEOREM 5: The Theorem follows from Theorem 2(b), Theorem 3(d)–(e), Assumptions 10 and 11, and the continuous mapping theorem. *Q.E.D.*

PROOF OF THEOREM 6: Part (a) holds by Theorem 1(a) and Assumptions 9(a), 9(b), and 12, because $\beta_{1*} = \beta_{10}$ for $\theta_0 \in \Theta_0$. Part (b) holds by part (a). Part (c) holds by part (a), Theorem 1(b), Assumption 12, and the continuous mapping theorem. Part (d) follows from part (c), the continuous mapping theorem, and Lemma 8(a). *Q.E.D.*

PROOF OF LEMMA 3: We make one-term Taylor expansions of $D_*\ell_T(\theta,\pi)$ about θ_0 element by element using Theorem 6 of E1. Stacking the expansions gives

$$D_*\ell_T(\theta,\pi) = D_*\ell_T(\theta_0,\pi) + D_*^{\,2}\ell_T(\theta_\pi^\dagger,\pi)(\theta-\theta_0), \tag{9.5}$$

where $\theta_\pi^\dagger$ lies between θ and θ_0 and $\theta_\pi^\dagger$ may differ across the rows of $D_*^{\,2}\ell_T(\theta_\pi^\dagger,\pi)$. Thus, (7.1) holds with

$$B_{*T}^{-1\prime} R_T^*(\theta,\pi) = [B_{*T}^{-1\prime}(D_*^{\,2}\ell_T(\theta_\pi^\dagger,\pi) - D_*^{\,2}\ell_T(\theta_0,\pi))B_T^{-1}]B_T(\theta-\theta_0). \tag{9.6}$$

Thus, by Assumption 2^{2^*}(c), $\sup_{\theta\in\Theta:\|\theta-\theta_0\|\le\gamma_T}\|B_{*T}^{-1\prime}R_T^*(\theta,\pi)\|/\|B_T(\theta-\theta_0)\| = o_{p\pi}(1)$ for all $\gamma_T \to 0$. This implies Assumption 13(a) by taking $\gamma_T = \gamma/\lambda_{\min}(B_T)$. *Q.E.D.*

PROOF OF LEMMA 4: We establish part (a) as follows. By Assumptions 3 and 7(a), $\mathcal{J}_{*\psi T\bullet} \Rightarrow \mathcal{J}_{*\psi\bullet} = \mathbf{0}$. Hence, by the continuous mapping theorem,

$$\sup_{\pi\in\Pi}\|\mathcal{J}_{*\psi T\pi}\| \to^d 0 \qquad \text{and} \qquad \mathcal{J}_{*\psi T\pi} = \mathcal{J}_{\psi *T\pi}' = o_{p\pi}(1). \tag{9.7}$$

This implies that

$$Z_{T\pi} := \mathcal{J}_{T\pi}^{-1}B_T^{-1\prime}D\ell_T(\theta_0,\pi) = \begin{pmatrix} \mathcal{J}_{*T\pi}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathcal{J}_{\psi T\pi}^{-1}\end{pmatrix} B_T^{-1\prime}D\ell_T(\theta_0,\pi) + o_{p\pi}(1) \qquad \text{and} \tag{9.8}$$

$$Z_{*T\pi} = \mathcal{J}_{*T\pi}^{-1}B_{*T}^{-1\prime}D_*\ell_T(\theta_0,\pi) + o_{p\pi}(1),$$

which establishes the second equality of part (a) because $Z_{\beta_1 T\pi} := H_1 Z_{*T\pi}$.

Next, by Assumptions 3, 4_0, and 13(a),

$$\begin{aligned}
&B_{*T}^{-1\prime}D_*\ell_T(\hat\theta_{0\pi},\pi)\\
&\quad = B_{*T}^{-1\prime}D_*\ell_T(\theta_0,\pi) + [\mathcal{J}_{*T\pi} \vdots \mathcal{J}_{*\psi T\pi}]B_T(\hat\theta_{0\pi}-\theta_0) + B_{*T}^{-1\prime}R_T^*(\hat\theta_{0\pi},\pi)\\
&\quad = O_{p\pi}(1), \qquad \text{because}
\end{aligned}$$

$$\begin{aligned}
&\|B_{*T}^{-1\prime}R_T^*(\hat\theta_{0\pi},\pi)\|1(\|B_T(\hat\theta_{0\pi}-\theta_0)\|\le\gamma)\\
&\quad \le \sup_{\theta\in\Theta_0:\|B_T(\theta-\theta_0)\|\le\gamma}\|B_{*T}^{-1\prime}R_T^*(\theta,\pi)\| = o_{p\pi}(1)
\end{aligned} \tag{9.9}$$

for all $0<\gamma<\infty$ and for any $\varepsilon>0$ there exists a $\gamma<\infty$ such that $\limsup_{T\to\infty}P(\|B_T(\hat\theta_{0\pi}-\theta_0)\|>\gamma) < \varepsilon$. By Assumption 9(a), $\hat\beta_{10\pi} = \beta_{1*} = \beta_{10}$. This, (9.7)–(9.9), and Assumptions 9(b) and 13(b) give

$$\begin{aligned}
&H_1\hat{\mathcal{J}}_{*T\pi}^{-1}B_{*T}^{-1\prime}D_*\ell_T(\hat\theta_{0\pi},\pi)\\
&\quad = H_1\mathcal{J}_{*T\pi}^{-1}B_{*T}^{-1\prime}D_*\ell_T(\hat\theta_{0\pi},\pi) + o_{p\pi}(1)\\
&\quad = H_1\mathcal{J}_{*T\pi}^{-1}B_{*T}^{-1\prime}D_*\ell_T(\theta_0,\pi) + H_1[I_{p+q} : \mathcal{J}_{*T\pi}^{-1}\mathcal{J}_{*\psi T\pi}]B_T(\hat\theta_{0\pi}-\theta_0) + o_{p\pi}(1)\\
&\quad = H_1\mathcal{J}_{*T\pi}^{-1}B_{*T}^{-1\prime}D_*\ell_T(\theta_0,\pi) + o_{p\pi}(1),
\end{aligned} \tag{9.10}$$

which establishes the first equality of part (a).

For part (b), let $\gamma_{T\pi} := (H_1\hat{\mathcal{J}}_{*T\pi}^{-1}H_1')^{-1/2}\,ds_\pi$ and $\xi_{T\pi} := (H_1\hat{\mathcal{J}}_{*T\pi}^{-1}H_1')^{-1/2}H_1\hat{\mathcal{J}}_{*T\pi}^{-1}B_{*T}^{-1\prime}\cdot D_*\ell_T(\hat\theta_{0\pi},\pi)$. By part (a) and Assumptions 3 and 13(b), $\xi_{T\pi} = O_{p\pi}(1)$. Thus,

$$\|\gamma_{T\pi}-\xi_{T\pi}\|^2 := \hat q_{\beta_1 T}(ds_\pi,\pi) \le \hat q_{\beta_1 T}(0,\pi) + o_{p\pi}(1) = \xi_{T\pi}'\xi_{T\pi} + o_{p\pi}(1) = O_{p\pi}(1). \tag{9.11}$$

Hence, $\gamma_{T\pi} = O_{p\pi}(1)$. This and Assumptions 3 and 13(b) yield part (b).

Part (c) follows straightforwardly from Assumption 13(b), the definitions of $\hat{q}_{\beta_1 T}(\lambda_{\beta_1}, \pi)$ and $q_{\beta_1 T}(\lambda_{\beta_1}, \pi)$, and parts (a) and (b).

For part (d), define $\tilde{ds}_\pi$ to satisfy $\tilde{ds}_\pi \in B_{\beta_1 T}(\mathscr{B}_1 - \beta_{1*})$ and $q_{\beta_1 T}(\tilde{ds}_\pi, \pi) = \inf_{\lambda_{\beta_1} \in B_{\beta_1 T}(\mathscr{B}_1 - \beta_{1*})} q_{\beta_1 T}(\lambda_{\beta_1}, \pi) + o_{p\pi}(1)$. By the method of proof of part (b), we obtain $\tilde{ds}_\pi = O_{p\pi}(1)$. As in the proof of part (c), we then obtain $q_{\beta_1 T}(\tilde{ds}_\pi, \pi) = \hat{q}_{\beta_1 T}(\tilde{ds}_\pi, \pi) + o_{p\pi}(1)$ using part (a). This result, part (c), and the definitions of ds_π and $\tilde{ds}_\pi$ give

$$(9.12) \qquad \begin{aligned} o_{p\pi}(1) &\le \hat{q}_{\beta_1 T}(\tilde{ds}_\pi, \pi) - \hat{q}_{\beta_1 T}(ds_\pi, \pi) \\ &= q_{\beta_1 T}(\tilde{ds}_\pi, \pi) - q_{\beta_1 T}(ds_\pi, \pi) + o_{p\pi}(1) \le o_{p\pi}(1). \end{aligned}$$

Equation (9.12) and the definition of $\tilde{ds}_\pi$ establish part (d).

Part (e) holds by the proof of Lemma 2 of E1 with θ, λ, B_T, Z_T, $\mathscr{J}_T$, $q_T(\lambda)$, Θ, Λ, and $o_p(1)$ replaced by β_1, λ_{β_1}, $B_{\beta_1 T}$, $Z_{\beta_1 T\pi}$, $(H_1 \mathscr{J}_{*T\pi}^{-1} H_1')^{-1}$, $q_{\beta_1 T}(\lambda_{\beta_1}, \pi)$, $\mathscr{B}_1$, Λ_{β_1}, and $o_{p\pi}(1)$ respectively. We use the fact that Assumptions 5, 7(b), and 9(c) imply that Assumption 5 holds with Θ and Λ replaced by $\mathscr{B}_1$ and Λ_{β_1}.

Part (f) holds by the proof of Theorem 4(a) of E1 with $\hat{\lambda}_\beta$, Z, $\mathscr{J}$, G, G_*, G_δ, $q_\beta(\lambda_\beta), \ldots$ replaced by $\hat{\lambda}_{\beta_1 T\pi}$, $Z_{T\pi}$, $\mathscr{J}_{T\pi}$, $G_{T\pi}$, $G_{*T\pi}$, $G_{\delta T\pi}$, $q_{\beta_1 T}(\lambda_{\beta_1}, \pi), \ldots$, where $G_{T\pi} := B_T^{-1'} D\ell_T(\theta_0, \pi)$ and $G_{T\pi} := (G'_{*T\pi}, G'_{\psi T\pi})' := (G'_{\beta_1 T\pi}, G'_{\beta_2 T\pi}, G'_{\delta T\pi}, G'_{\psi T\pi})'$. In place of Assumption 7(a) of E1, which assumes that $\mathscr{J}_{*\psi} = \mathscr{J}'_{\psi *} = \mathbf{0}$, we use (9.7), which yields $\mathscr{J}_{*\psi T\pi} = o_{p\pi}(1)$. In consequence, some of the equalities in the proof of Theorem 4(a) of E1 are equalities only up to $o_{p\pi}(1)$ terms.

Part (g) holds by the proof of Theorem 3(a) of E1 with θ, $\hat{\theta}$, λ, $\hat{\lambda}_T$, Λ, B_T, Z_T, $\mathscr{J}_T$, $o_p(1), \ldots$ replaced by $\beta_1, \hat{\beta}_{1\pi}, \lambda_{\beta_1}, \hat{\lambda}_{\beta_1 T\pi}, \Lambda_{\beta_1}, B_{\beta_1 T}, Z_{\beta_1 T\pi}, (H_1 \mathscr{J}_{*T\pi}^{-1} H_1')^{-1}$, $o_{p\pi}(1), \ldots$ respectively. The proof uses the fact that $q_{\beta_1 T}(ds_\pi, \pi) = q_{\beta_1 T}(\hat{\lambda}_{\beta_1 T\pi}, \pi) + o_{p\pi}(1)$, which holds by parts (d)–(f) of the Lemma. *Q.E.D.*

PROOF OF THEOREM 7: Parts (a) and (b) follow from Lemma 4(g) and Assumption 12. Part (c) follows from part (a), Assumption 12, Theorem 1(b), and the continuous mapping theorem (CMT). Part (d) follows from part (c) and the CMT. *Q.E.D.*

PROOF OF LEMMA 5: Define the $p+q$ vector $r_T(\theta, \pi)$ via

$$(9.13) \qquad \begin{aligned} &T^{-1} D_* \ell_T(\theta, \pi) - D_* \ell(\theta, \pi) \\ &\quad := T^{-1} D_* \ell_T(\theta_0, \pi) - D_* \ell(\theta_0, \pi) \\ &\qquad + (T^{-1} D_*{}^2 \ell_T(\theta_0, \pi) - D_*{}^2 \ell(\theta_0, \pi))(\theta - \theta_0) + r_T(\theta, \pi). \end{aligned}$$

By the stochastic differentiability Assumption 13*(e),

$$(9.14) \qquad \sup_{\theta \in \Theta : \|\theta - \theta_0\| \le \gamma_T} T \|r_T(\theta, \pi)\| / (1 + \|T^{1/2}(\theta - \theta_0)\|)^2 = o_{p\pi}(1)$$

for all $\gamma_T \to 0$.

By Theorem 6 of E1 and Assumptions 13*(c) and 13*(d), element by element one-term Taylor expansions of $D_* \ell(\theta, \pi)$ about θ_0 stacked give: $\forall \theta \in \Theta \cap C(\theta_0, \varepsilon)$,

$$(9.15) \qquad D_* \ell(\theta, \pi) = D_* \ell(\theta_0, \pi) + D_*{}^2 \ell(\theta_\pi^\dagger, \pi)(\theta - \theta_0),$$

where $\theta_\pi^\dagger$ lies between θ and θ_0 and may differ across the rows of $D_*{}^2 \ell(\theta_\pi^\dagger, \pi)$.

Combining (9.13) and (9.15) gives

$$(9.16) \qquad \begin{aligned} &T^{-1} D_* \ell_T(\theta, \pi) \\ &\quad = T^{-1} D_* \ell_T(\theta_0, \pi) + T^{-1} D_*{}^2 \ell_T(\theta_0, \pi)(\theta - \theta_0) \\ &\qquad + (D_*{}^2 \ell(\theta_\pi^\dagger, \pi) - D_*{}^2 \ell(\theta_0, \pi))(\theta - \theta_0) + r_T(\theta, \pi) \\ &\quad = T^{-1} D_* \ell_T(\theta_0, \pi) + T^{-1} D_*{}^2 \ell_T(\theta_0, \pi)(\theta - \theta_0) + o(\|\theta - \theta_0\|) + r_T(\theta, \pi), \end{aligned}$$

where the second equality uses the continuity of $D_*{}^2 \ell(\theta, \pi)$ at θ_0 uniformly over Π.

Combining (9.16) with (7.1) divided by T gives

$$T^{-1}R_T^*(\theta,\pi)=o(\|\theta-\theta_0\|)+r_T(\theta,\pi). \tag{9.17}$$

This result and (9.14) imply Assumption 13(a) by taking $\gamma_T=\gamma/\lambda_{\min}(B_T)$. *Q.E.D.*

D. Proofs for the GARCH Example

To verify Assumptions 1* and 2*, we show that $\ell_T(\theta,\pi)$ is closely approximated by the sum of stationary and ergodic terms $\mathscr{L}_T(\theta,\pi)=\sum_{t=1}^T \ell_{tt}(\theta,\pi)$. Note that we can write $\ell_T(\theta,\pi)=\sum_{t=1}^T \ell_t^*(\theta,\pi)$, where $\ell_t^*(\theta,\pi)=-\frac{1}{2}\ln(2\tilde{\pi})-\frac{1}{2}\ln(h_t^*(\theta,\pi))-\frac{1}{2}e_t^2(\theta)/h_t^*(\theta,\pi)$. First, we show that

$$\sum_{t=2}^{\infty}\sup_{\theta\in\Theta,\pi\in\Pi}|d_t(\theta,\pi)|<\infty \quad \text{a.s.,} \quad \text{where} \quad d_t(\theta,\pi)=\ell_{tt}(\theta,\pi)-\ell_t^*(\theta,\pi). \tag{9.18}$$

Some calculations show that for $t\geq 2$,

$$\begin{aligned}
&h_t(\theta,\pi)=h_t^*(\theta,\pi)+\pi^{t-1}A(\theta,\pi), \qquad \text{where}\\
&A(\theta,\pi):=\left(\delta+\beta_1\sum_{k=1}^{\infty}\pi^k e_{-k}^2(\theta)-h_1^*(\theta,\pi)\right),\\
&2|d_t(\theta,\pi)|=|\ln(h_t(\theta,\pi)/h_t^*(\theta,\pi))+e_t^2(\theta)/h_t(\theta,\pi)-e_t^2(\theta)/h_t^*(\theta,\pi)|\\
&\qquad\leq \ln\left(1+\pi_u^{t-1}\frac{|A(\theta,\pi)|}{\delta_\ell(1-\pi_u)}\right)+e_t^2(\theta)\left|\frac{\pi_u^{t-1}A(\theta,\pi)}{h_t(\theta,\pi)h_t^*(\theta,\pi)}\right|\\
&\qquad\leq \pi_u^{t-1}\frac{|A(\theta,\pi)|}{\delta_\ell(1-\pi_u)}(1+e_t^2(\theta)/\delta_\ell), \qquad \text{and}\\
&2\sum_{t=2}^{\infty}\sup_{\theta\in\Theta,\pi\in\Pi}|d_t(\theta,\pi)|\leq\frac{\sup_{\theta\in\Theta,\pi\in\Pi}|A(\theta,\pi)|}{\delta_\ell(1-\pi_u)}\sum_{t=2}^{\infty}\pi_u^{t-1}\left(1+\sup_{\theta\in\Theta}e_t^2(\theta)/\delta_\ell\right).
\end{aligned} \tag{9.19}$$

The right-hand side of the last expression in (9.19) is finite with probability one, because (i) the fact that $E\sup_{\theta\in\Theta}e_t^2(\theta)$ is finite and does not depend on t implies that $E\sum_{k=1}^{\infty}\pi_u^k\sup_{\theta\in\Theta}e_{-k}^2(\theta)<\infty$, which, in turn, implies that $\sum_{k=1}^{\infty}\pi_u^k\sup_{\theta\in\Theta}e_{-k}^2(\theta)<\infty$ a.s., (ii) the assumption that $\sup_{\theta\in\Theta,\pi\in\Pi}h_1^*(\theta,\pi)<\infty$ and result (i) imply that $\sup_{\theta\in\Theta,\pi\in\Pi}|A(\theta,\pi)|<\infty$ a.s., and (iii) result (i) implies that $\sum_{t=2}^{\infty}\pi_u^{t-1}(1+\sup_{\theta\in\Theta}e_t^2(\theta)/\delta_\ell)<\infty$ a.s. Hence, (9.18) holds.

Equation (9.18) and Kronecker's Lemma imply that

$$\sup_{\theta\in\Theta,\pi\in\Pi}T^{-1}|\mathscr{L}_T(\theta,\pi)-\ell_T(\theta,\pi)|\leq T^{-1}\sum_{t=1}^{T}\sup_{\theta\in\Theta,\pi\in\Pi}|d_t(\theta,\pi)|\to^p 0. \tag{9.20}$$

Now, we verify Assumption 1*. Given (9.20), it suffices to verify Assumption 1*(a) with $T^{-1}\mathscr{L}_T(\theta,\pi)$ in place of $T^{-1}\ell_T(\theta,\pi)$. To do so, we use the uniform law of large numbers given in Theorem 6 of Andrews (1992) employing Assumption TSE-1D. This uniform law of large numbers holds, because $\{\ell_{tt}(\theta,\pi):t=\ldots,1,2,\ldots\}$ is stationary and ergodic, $\ell_{tt}(\theta,\pi)$ is continuous in θ and π a.s., and

$$\begin{aligned}
E\sup_{\theta\in\Theta,\pi\in\Pi}|2\ell_{tt}(\theta,\pi)|&\leq\ln(2\tilde{\pi})+\ln(\delta_\ell)+E\ln\left(\delta_u+\beta_{1u}\sum_{k=1}^{\infty}\pi_u^k\sup_{\theta\in\Theta}e_{t-k-1}^2(\theta)\right)\\
&\quad+E\sup_{\theta\in\Theta}e_t^2(\theta)/\delta_\ell<\infty.
\end{aligned} \tag{9.21}$$

The limit function $\ell(\theta,\pi)$ of Assumption 1*(a) equals $E\ell_{tt}(\theta,\pi)$.

Next, we verify Assumption 1*(b). Because $\beta_{10}=0$, $\ell(\theta_0,\pi)$ does not depend on π. The uniform LLN used above delivers continuity of the limit function $\ell(\theta,\pi) := E\ell_{tt}(\theta,\pi)$ on the compact set $\Theta\times\Pi$. In consequence, it suffices to show that $\ell(\theta,\pi)$ is uniquely maximized over Θ at θ_0 for each $\pi\in\Pi$. We have

$$(9.22)\qquad E\ell_{tt}(\theta,\pi) = -\frac{1}{2}\ln(2\tilde{\pi}) - \frac{1}{2}E\ln(h_t(\theta,\pi)) - \frac{1}{2}Ee_t^2(\theta)/h_t(\theta,\pi) \qquad \text{and}$$

$$\begin{aligned} Ee_t^2(\theta)/h_t(\theta,\pi) &= E\varepsilon_t^2/h_t(\theta,\pi) + (\psi-\psi_0)'EX_tX_t'(\psi-\psi_0)\\ &\geq E\varepsilon_t^2/h_t(\theta,\pi)\\ &= Eh_{0t}/h_t(\theta,\pi)\end{aligned}$$

with strict inequality unless $\psi=\psi_0$, because EX_tX_t' is positive definite (pd). The function $\ln(x)+y/x$ is uniquely minimized over $x>0$ at $x=y$. Hence,

$$(9.23)\qquad E\ell_{tt}(\theta,\pi) \leq -\frac{1}{2}\ln(2\tilde{\pi}) - \frac{1}{2}E\ln(h_{t0}) - \frac{1}{2} = \ell(\theta_0,\pi)$$

with equality iff $\psi=\psi_0$ and $P(h_t(\theta,\pi)=h_{0t})=1$.

Hence, Assumption 1*(b) holds provided for any θ with $\psi=\psi_0$, $P(h_t(\theta,\pi)=h_{0t})=1$ iff $\theta=\theta_0$. For θ with $\psi=\psi_0$, we have

$$(9.24)\qquad h_t(\theta,\pi)-h_{0t} = \pi(h_{t-1}(\theta,\pi)-h_{0t-1}) + (\delta_0 z_{t-1}^2, 1-\pi)'(\bar{\theta}-\bar{\theta}_0),$$

where $\bar{\theta}=(\beta_1,\delta)'$ and $\bar{\theta}_0=(0,\delta_0)'$. By stationarity of $\{h_t(\theta,\pi)-h_{0t} : t=\ldots,0,1\ldots\}$, $h_t(\theta,\pi)-h_{0t}=0$ a.s. iff $h_{t-1}(\theta,\pi)-h_{0t-1}=0$ a.s. Combining this with (9.24), we find that $h_t(\theta,\pi)-h_{0t}=0$ a.s. iff $(\delta_0 z_{t-1}^2, 1-\pi)'(\bar{\theta}-\bar{\theta}_0)=0$ a.s. Because $\delta_0>0$ and z_t is not a constant by (2.3), the latter holds only if $\beta_1=0$. Given $\beta_1=0$, $(\delta_0 z_{t-1}^2, 1-\pi)'(\bar{\theta}-\bar{\theta}_0)=0$ a.s. implies that $\delta=\delta_0$, because $\pi<1$. This completes the verification of Assumption 1*(b).

Next, we verify Assumption 2* for $\ell_T(\theta)$ by showing that it holds for $\mathscr{L}_T(\theta)$ and that $\ell_T(\theta)$ is closely approximated by $\mathscr{L}_T(\theta)$ in the sense that (9.1) holds. To establish (9.1), we note that the left-hand side of (9.1) is less than or equal to

$$(9.25)\qquad 2\sum_{t=1}^{T}\sup_{\theta\in\Theta:\|\theta-\theta_0\|\leq\gamma_T,\,\pi\in\Pi}|d_t(\theta,\pi)|.$$

The function $d_t(\theta,\pi)$ is continuous in θ uniformly over $\pi\in\Pi$ a.s. Hence, $\sup\{|d_t(\theta,\pi)| : \theta\in\Theta, \|\theta-\theta_0\|\leq\gamma_T,\ \pi\in\Pi\}\to 0$ a.s. as $T\to\infty$ $\forall t\geq 1$. By (9.18), given $\varepsilon>0$, $\exists T_1<\infty$ such that $\sum_{t=T_1+1}^{\infty}\sup_{\theta\in\Theta,\pi\in\Pi}|d_t(\theta,\pi)|<\varepsilon/4$. And, given $\varepsilon>0$ and $T_1<\infty$, $\exists T_2<\infty$ such that $\forall T\geq T_2$ we have $\sup_{\theta\in\Theta:\|\theta-\theta_0\|\leq\gamma_T,\pi\in\Pi}|d_t(\theta,\pi)|<\varepsilon/(4T_1)$ $\forall t\leq T_1$. Combining these results, we find that the expression in (9.25) is less than or equal to

$$(9.26)\qquad 2\sum_{t=1}^{T_1}\sup_{\theta\in\Theta:\|\theta-\theta_0\|\leq\gamma_T,\,\pi\in\Pi}|d_t(\theta,\pi)| + 2\sum_{t=T_1+1}^{\infty}\sup_{\theta\in\Theta,\,\pi\in\Pi}|d_t(\theta,\pi)| < \varepsilon.$$

Hence, (9.1) holds.

We verify Assumption 2* for $\mathscr{L}_T(\theta)$ using Assumption 2^{2^*} and the comment following it. (Note that the latter applies because it follows from the result in the next paragraph that $-T^{-1}\partial^2\mathscr{L}_T(\theta_0,\pi)/\partial\theta\,\partial\theta' = \mathscr{J}_\pi + o_{p\pi}(1)$.) Assumption 2^{2^*}(a) holds with $\Theta^+ = \Theta\cap C(\theta_0,\varepsilon)$ for some $\varepsilon>0$ small, because $\Theta-\theta_0$ is a union of orthants local to **0**. Assumption 2^{2^*}(b) holds straightforwardly.

We verify Assumption 2^{2^*}(c) for $\mathscr{L}_T(\theta,\pi)$ by showing that

$$\sup_{\theta\in\Theta_0,\,\pi\in\Pi}\left\|\frac{\partial^2}{\partial\theta\,\partial\theta'}\mathscr{L}_T(\theta,\pi) - E\frac{\partial^2}{\partial\theta\,\partial\theta'}\ell_{tt}(\theta,\pi)\right\|\to_p 0,$$

for some set $\Theta_0 \subset \Theta$ that contains $\Theta \cap S(\theta_0, \varepsilon)$ for some $\varepsilon > 0$, and $E\partial^2 \ell_{tt}(\theta)/\partial\theta\,\partial\theta'$ is continuous at θ_0. Both of these results follow from the uniform LLN given in Theorem 6 of Andrews (1992) using Assumption TSE-1D provided

$$(9.27)\qquad E \sup_{\theta \in \Theta_0, \pi \in \Pi} \left\| \frac{\partial^2}{\partial\theta\,\partial\theta'} \ell_{tt}(\theta, \pi) \right\| < \infty,$$

because $(\partial^2/\partial\theta\,\partial\theta')\ell_{tt}(\theta,\pi)$ is stationary and ergodic and continuous in (θ,π) a.s.

Let h, e, ∂, ∂', and ∂^2 abbreviate $h_t(\theta,\pi)$, $e_t(\theta)$, $\partial/\partial\theta$, $\partial/\partial\theta'$, and $\partial^2/\partial\theta\,\partial\theta'$ respectively. Some calculations show that

$$(9.28)\qquad 2\partial^2 \ell_{tt}(\theta,\pi) = (h^{-1}e^2 - 1)(h^{-1}\partial^2 h - h^{-2}(\partial h)^2) + h^{-2}e\partial h(2\partial' e - h^{-1}e\partial' h)$$
$$- 2h^{-1}e\partial^2 e - 2h^{-1}\partial e(\partial' e - h^{-1}e\partial' h),$$
$$h^{-1} \le \delta_\ell^{-1}, \qquad \|\partial e\| \le \|X_t\|, \qquad \partial^2 e = 0, \qquad e \le C(|z_t| + \|X_t\|), \qquad \text{and}$$
$$\|\partial h\| + \|\partial^2 h\| \le C \sum_{k=0}^{\infty} \pi_u^k (1 + |z_{t-k-1}|^2 + \|X_{t-k-1}\|^2),$$

for some constant $C < \infty$ that does not depend on θ or π. In consequence,

$$(9.29)\qquad \|\partial^2 \ell_{tt}(\theta,\pi)\| \le C(1 + |z_t|^2 + \|X_t\|^2)\left(\sum_{k=0}^{\infty} \pi_u^k (1 + |z_{t-k-1}|^2 + \|X_{t-k-1}\|^2)\right)^2$$
$$+ C(|z_t| + \|X_t\|)\|X_t\| \sum_{k=0}^{\infty} \pi_u^k (1 + |z_{t-k-1}|^2 + \|X_{t-k-1}\|^2) + C\|X_t\|^2.$$

Hence, $E \sup_{\theta \in \Theta_0, \pi \in \Pi} \|\partial^2 \ell_{tt}(\theta,\pi)\| < \infty$ by the moment conditions of (2.3).

Next, we verify Assumption 3*. By Theorem 10.2 of Pollard (1990), $B_T^{-1'} D\ell_T(\theta_0, \bullet) \Rightarrow G_\bullet$ if (i) Π is totally bounded, (ii) the finite dimensional distributions of $B_T^{-1} D\ell_T(\theta_0,\bullet)$ converge to those of $G_\bullet$, and (iii) $\{B_T^{-1'} D\ell_T(\theta_0,\bullet) : T \ge 1\}$ is stochastically equicontinuous. Condition (i) holds because $\Pi \subset [0,1]$. The variable $B_T^{-1'} \times D\ell_T(\theta_0,\pi)$ equals $T^{-1/2}\sum_{t=1}^T \partial\ell_{tt}(\theta_0,\pi)$, which is a normalized sample average of strictly stationary martingale difference random variables. Using the definition of $\partial\ell_{tt}(\theta_0,\pi)$ in (4.3) and the moment conditions in (2.3), we obtain $E \sup_{\pi \in \Pi} \|\partial\ell_{tt}(\theta_0,\pi)\|^2 < \infty$. In consequence, the martingale difference central limit theorem of Billingsley (1968, Thm. 23.1) implies that each of the finite dimensional distributions of $B_T^{-1'} D\ell_T(\theta_0,\bullet)$ converges in distribution to a multivariate normal distribution with covariance given in (4.4), which by definition equals the covariance of $G_\bullet$. Thus, condition (ii) holds. Stochastic equicontinuity of $\{T^{-1/2}\sum_{t=1}^T \partial\ell_{tt}(\theta_0,\bullet) : T \ge 1\}$ is established by the same argument as in Andrews and Ploberger (1996, p. 1340, (A.14) and the following paragraph) for $\nu_T^*(\pi)$ with Y_t and Y_{t-k-1} replaced by $\delta_0^{1/2}(z_t^2 + 1)$ and z_{t-k-1} respectively. Hence, condition (iii) holds.

The matrix $\mathcal{J}_\pi$ is symmetric and nonrandom, so, to verify Assumption 3*, it remains to show that $\sup_{\pi \in \Pi} \lambda_{\max}(\mathcal{J}_\pi) < \infty$ and $\inf_{\pi \in \Pi} \lambda_{\min}(\mathcal{J}_\pi) > 0$. From (4.3),

$$(9.30)\qquad \det(\mathcal{J}_\pi) = \det\left(\frac{Ez_t^4 - 1}{4\delta_0^2(1 - \pi^2)}\right) \times \det(\delta_0^{-1} EX_t X_t'),$$

which yields the desired results because $0 < Ez_t^4 - 1 < \infty$, $\sup\{\pi^2 : \pi \in \Pi\} < 1$, $\delta_0 > 0$, and $EX_t X_t'$ is positive definite by (2.3).

E. Proofs for the Random Coefficient Example

First, we verify Assumption 1*. Assumption 1*(a) holds by the uniform LLN of Andrews (1992, Theorem 4) using Assumption TSE-1D and the standard pointwise LLN for iid random variables with finite mean. The function $\ell(\theta,\pi)$ of Assumption 1*(a) is

$$(9.31)\qquad \ell(\theta,\pi) := -\frac{1}{2}\ln(2\tilde{\pi}) - \frac{1}{2}E\ln(\delta_2 + X_t'\Omega(\beta_1,\delta_1,\pi)X_t) - \frac{1}{2}E(Y_t-\psi_2-X_t'\psi_1)^2/(\delta_2+X_t'\Omega(\beta_1,\delta_1,\pi)X_t).$$

Sufficient conditions for Assumption 1*(b) are: (i) $\ell(\theta,\pi)<\ell(\theta_0)$ for all $\theta\in\Theta$ with $\theta\neq\theta_0$ and all $\pi\in\Pi$, (ii) $\ell(\theta,\pi)$ is continuous on $\Theta\times\Pi$, and (iii) $\Theta\times\Pi$ is compact. Condition (ii) holds straightforwardly. Condition (iii) holds by assumption.

To show condition (i), we first show that for any $(\beta_1,\delta_1,\delta_2)$ in the parameter space the third summand of $\ell(\theta,\pi)$ is uniquely maximized by $(\psi_1,\psi_2)=(\psi_{10},\psi_{20})$. The third summand of $\ell(\theta,\pi)$ can be written as the sum of $-\frac{1}{2}E(\psi_2-\psi_{20}+X_t'(\psi_1-\psi_{10}))^2/(\delta_2+X_t'\Omega(\beta_1,\delta_1,\pi)X_t)$ and a term that does not involve (ψ_1,ψ_2). In consequence, the third summand is uniquely maximized at (ψ_{10},ψ_{20}) if $E(a'W_t)^2/(\delta_2+X_t'\Omega(\beta_1,\delta_1,\pi)X_t)>0$ whenever $a\neq 0$ or, equivalently, if $E(a'W_t)^2>0$ whenever $a\neq 0$ or, equivalently, if $EW_tW_t'>0$, where the second "if" holds because $1/(\delta_2+X_t'\Omega(\beta_1,\delta_1,\pi)X_t)$ is positive with probability one. The last condition holds by (4.14).

Next, we show that, for any parameter $\theta=(\beta_1,\delta_1',\delta_2,\psi_{10}',\psi_{20})'$ with $(\beta_1,\delta_1',\delta_2)'\neq(0,\delta_{10}',\delta_{20})'$, $\ell(\theta,\pi)<\ell(\theta_0)$ for all $\pi\in\Pi$. For θ as above, $\ell(\theta,\pi)$ can be written as

$$(9.32)\qquad \ell(\theta,\pi) = -\frac{1}{2}\ln(2\tilde{\pi}) - \frac{1}{2}E\ln(\delta_2+X_t'\Omega(\beta_1,\delta_1,\pi)X_t) + \frac{1}{2}E\frac{\delta_{20}+X_t'\Omega(\beta_1,\delta_{10},\pi)X_t}{\delta_2+X_t'\Omega(\beta_1,\delta_1,\pi)X_t}.$$

The function $(\ln x)+y/x$ is uniquely minimized over $x\in R$ at $x=y$. Thus, $\ell(\theta,\pi)<\ell(\theta_0,\pi)$ unless

$$(9.33)\qquad P(\delta_2+X_t'\Omega(\beta_1,\delta_1,\pi)X_t-(\delta_{20}+X_t'\Omega(0,\delta_{10},\pi)X_t)=0)=1.$$

By the form of $\Omega(\beta_1,\delta_1,\pi)$, $\Omega(\beta_1,\delta_1,\pi)=\Omega(0,\delta_{10},\pi)$ only if $(\beta_1,\delta_1)=(0,\delta_{10})$. In consequence, the left-hand side of the equation in the probability in (9.33) is of the form $c'\tilde{X}_t$ for a vector $c\neq\mathbf{0}$ because $(\beta_1,\delta_1,\delta_2)\neq(0,\delta_{10},\delta_{20})$. But, $E\tilde{X}_t\tilde{X}_t'>0$ (by (4.14)) implies that $P(a'\tilde{X}_t=0)<1$ for all $a\neq\mathbf{0}$. Thus, (9.33) cannot hold, condition (i) does hold, and the verification of Assumption 1* is complete.

We verify Assumption 2* for the random coefficient example with $D\ell_T(\theta_0,\pi)$ and $D^2\ell_T(\theta_0,\pi)$ of (3.3) as defined in (4.12) using Assumption 2^{2^*} and the remark following Assumption 2^{2^*}. The proof is analogous to that for the random coefficient example of E1 given in Section 3.6.1 of E1. In the present case, the matrix of ℓ/r partial derivatives of order two of $\ell_T(\theta,\pi)$ is

$$(9.34)\qquad \frac{\partial^2}{\partial\theta\,\partial\theta'}\ell_T(\theta,\pi) := -\sum_{t=1}^{T}\begin{pmatrix} \frac{2\,\mathrm{res}_t^2(\theta)-\mathrm{var}_t(\theta,\pi)}{\mathrm{var}_t^3(\theta,\pi)}\tilde{W}_t(\theta,\pi)\tilde{W}_t(\theta,\pi)' & \frac{\mathrm{res}_t(\theta)}{\mathrm{var}_t^2(\theta,\pi)}W_t\tilde{W}_t(\theta,\pi)' \\ \frac{\mathrm{res}_t(\theta)}{\mathrm{var}_t^2(\theta,\pi)}\tilde{W}_t(\theta,\pi)W_t' & \frac{1}{\mathrm{var}_t(\theta,\pi)}W_tW_t' \end{pmatrix},$$

where $\tilde{W}_t(\theta,\pi):=(X_{1t}^*(\pi),X_{2t}^*(\delta_1)',1)'$.

Next, we verify Assumption 3*. Note that (4.14) implies that $EW_tW_t'/\mathrm{var}_t(\theta_0)>0$, $E\tilde{W}_t(\pi)\tilde{W}_t(\pi)'/\mathrm{var}_t^2(\theta_0)>0\ \forall\pi\in\Pi$, and $E(\mathrm{res}_t^2(\theta_0)-\mathrm{var}_t(\theta_0))^2\tilde{W}_t(\pi)\tilde{W}_t(\pi)'/\mathrm{var}_t^4(\theta_0)>0\ \forall\pi\in\Pi$. Continuity in π of the latter two terms and compactness of Π, then yields $\inf_{\pi\in\Pi}\lambda_{\min}(\cdot)$ of the latter two terms to be positive. In consequence, $\inf_{\pi\in\Pi}\lambda_{\min}(\mathcal{J}_\pi)>0$. By (4.12), $\mathcal{J}_\pi$ is symmetric. By (4.13), continuity in π of $\mathcal{J}_\pi$, and compactness of Π, we have $\sup_{\pi\in\Pi}\lambda_{\max}(\mathcal{J}_\pi)<\infty$.

Thus, Assumption 3* holds provided $T^{-1/2}D\ell_T(\theta_0,\bullet) \Rightarrow G_\bullet$. By Theorem 10.2 of Pollard (1990), $T^{-1/2}D\ell_T(\theta_0,\bullet) \Rightarrow G_\bullet$ if (i) Π is totally bounded, (ii) the finite dimensional distributions of $T^{-1/2}D\ell_T(\theta_0,\bullet)$ converge to those of $G_\bullet$, and (iii) $\{T^{-1/2}D\ell_T(\theta_0,\bullet): T \geq 1\}$ is stochastically equicontinuous. Condition (i) holds by compactness of Π. Condition (ii) holds by the CLT for iid mean zero finite variance random variables using the definition of $D\ell_T(\theta_0,\pi)$ in (4.12) and the moment assumptions of (4.13). To obtain condition (iii), we write

$$
\begin{aligned}
(9.35)\qquad & \|T^{-1/2}D\ell_T(\theta_0,\pi) - T^{-1/2}D\ell_T(\theta_0,\overline{\pi})\| \\
&= \left\| T^{-1/2}\sum_{t=1}^{T}\frac{\mathrm{res}_t^2(\theta_0)-\mathrm{var}_t(\theta_0)}{2\mathrm{var}_t^2(\theta_0)} X_{1t}'(\Omega_1(\pi)-\Omega_1(\overline{\pi}))X_{1t}\right\| \\
&\leq \sum_{i=1}^{b}\sum_{j=1}^{b}\|\Omega_1(\pi)_{ij}-\Omega_1(\overline{\pi})_{ij}\| T^{-1/2}\sum_{t=1}^{T}|\mathrm{res}_t^2(\theta_0)-\mathrm{var}_t(\theta_0)|\cdot\|X_{ti}X_{tj}\|/\delta_{20} \\
&= \sum_{i=1}^{b}\sum_{j=1}^{b}\|\Omega_1(\pi)_{ij}-\Omega_1(\overline{\pi})_{ij}\| O_p(1),
\end{aligned}
$$

where $\Omega_1(\pi)_{ij}$ denotes the (i,j) element of $\Omega_1(\pi)$ and X_{ti} denotes the ith element of X_t. This result, the continuity of $\Omega_1(\pi)_{ij}$ in π, and the compactness of Π yield stochastic equicontinuity of $T^{-1/2}D\ell_T(\theta_0,\bullet)$: given any $\eta > 0$, there exists $\lambda > 0$ such that

$$
(9.36)\qquad \overline{\lim}_{T\to\infty} P\left(\sup_{\pi,\overline{\pi}\in\Pi:\|\pi-\overline{\pi}\|<\lambda}\|T^{-1/2}D\ell_T(\theta_0,\pi) - T^{-1/2}D\ell_T(\theta_0,\overline{\pi})\|\right) < \eta.
$$

REFERENCES

ANDREWS, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

——— (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

——— (1994a): "Asymptotics for Semiparametric Econometric Models Via Stochastic Equicontinuity," *Econometrica*, 62, 43–72.

——— (1994b): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, *Vol. IV*, ed. by R. F. Engle and D. McFadden. New York: North-Holland.

——— (1996): "Admissibility of the Likelihood Ratio Test when the Parameter Space Is Restricted under the Alternative," *Econometrica*, 64, 705–718.

——— (1997): "Estimation when a Parameter Is on a Boundary: Part II," unpublished manuscript, Cowles Foundation, Yale University.

——— (1998): "Hypothesis Testing with a Restricted Parameter Space," *Journal of Econometrics*, 84, 155–199.

——— (1999): "Estimation when a Parameter Is on a Boundary," *Econometrica*, 67, 1341–1383.

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal Tests when a Nuisance Parameter Is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

——— (1995): "Admissibility of the Likelihood Ratio Test when a Nuisance Parameter Is Present Only under the Alternative," *Annals of Statistics*, 23, 1609–1629.

——— (1996): "Testing for Serial Correlation Against an ARMA(1, 1) Model," *Journal of the American Statistical Association*, 91, 1331–1342.

BARLOW, R. E., D. J. BARTHOLOMEW, J. N. BREMNER, AND H. D. BRUNK (1972): *Statistical Inference Under Order Restrictions*: *The Theory and Application of Isotonic Regression*. New York: Wiley.

BARTHOLOMEW, D. J. (1959): "A Test of Homogeneity of Ordered Alternatives, I," *Biometrika*, 46, 36–48.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: Wiley.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

BOLLERSLEV, T., R. F. ENGLE, AND D. NELSON (1994): "ARCH Models," Ch. 49 in *Handbook of Econometrics*, *Vol. IV*, ed. by R. F. Engle and D. McFadden. New York: North-Holland, 2959–3038.

BROOKS, R. D., AND M. L. KING (1994): "Hypothesis Testing of Varying Coefficient Regression Models: Procedures and Applications," *Pakistan Journal of Statistics*, 10, 301–357.

CHANT, D. (1974): "On Asymptotic Tests of Composite Hypotheses in Nonstandard Conditions," *Biometrika*, 61, 291–298.

CHERNOFF, H. (1954): "On the Distribution of the Likelihood Ratio," *Annals of Mathematical Statistics*, 54, 573–578.

DAVIES, R. B. (1977): "Hypothesis Testing when a Nuisance Parameter Is Present Only under the Alternative," *Biometrika*, 64, 247–254.

——— (1987): "Hypothesis Testing when a Nuisance Parameter Is Present Only under the Alternative," *Biometrika*, 74, 33–43.

GEYER, C. (1994): "On the Asymptotics of Constrained M-estimation," *Annals of Statistics*, 22, 1993–2010.

GOURIEROUX, C., A. HOLLY, AND A. MONFORT (1982): "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters," *Econometrica*, 50, 63–80.

GOURIEROUX, C., AND A. MONFORT (1989): *Statistique et Modeles Econometriques*, Vol. 2. Paris: Economica. English translation (1995): *Statistics and Econometric Models*, Vol. 2. Translated by Q. Vuong. Cambridge, U.K.: Cambridge University Press.

HANSEN, B. E. (1996): "Inference when a Nuisance Parameter Is Not Identified under the Null Hypothesis," *Econometrica*, 64, 413–430.

HAUSMAN, J. A., AND D. WISE (1978): "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46, 403–426.

HILDRETH, C., AND J. P. HOUCK (1968): "Some Estimators for a Linear Model with Random Coefficients," *Journal of the American Statistical Association*, 63, 584–595.

JEGANATHAN, P. (1982): "On the Asymptotic Theory of Estimation When the Limit of the Loglikelihood Is Mixed Normal," *Sankhya*, Series A, 44, Part 2, 172–212.

KING, M. L., AND T. S. SHIVELY (1993): "Locally Optimal Testing When a Nuisance Parameter Is Present Only under the Alternative," *Review of Economics and Statistics*, 75, 1–7.

KING, M. L., AND P. X. WU (1997): "Locally Optimal One-sided Tests for Multiparameter Hypotheses," *Econometric Reviews*, 16, 131–156.

LE CAM, L. (1960): "Locally Asymptotically Normal Families of Distributions," *University of California Publications in Statistics*, 3, 37–98.

MCFADDEN, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995–1026.

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, *Vol. IV*, ed. by R. F. Engle and D. McFadden. New York: North-Holland.

PAKES, A., AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

PERLMAN, M. D. (1969): "One-sided Problems in Multivariate Analysis," *Annals of Mathematical Statistics*, 40, 549–567. Corrections in *Annals of Mathematical Statistics*, 42, 1777.

POLLARD, D. (1984): *Convergence of Stochastic Processes*. New York: Springer-Verlag.

——— (1985): "New Ways to Prove Central Limit Theorems," *Econometric Theory*, 1, 295–314.

——— (1990): *Empirical Processes*: *Theory and Applications*, CBMS Conference Series in Statistics and Probability, Vol. 2. Hayward, CA: Institute of Mathematical Statistics.

ROSENBERG, B. (1973): "The Analysis of a Cross-section of Time Series by Stochastically Convergent Parameter Regressions," *Annals of Social and Economic Measurement*, 2, 399–428.

SELF, S. G., AND K.-Y. LIANG (1987): "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions," *Journal of the American Statistical Association*, 82, 605–610.

SHAPIRO, A. (1985): "Asymptotic Distribution of Test Statistics in the Analysis of Moment Structures under Inequality Constraints, *Biometrika*, 72, 133–144.

VAN DER VAART, A. W., AND J. WELLNER (1996): *Weak Convergence and Empirical Processes*. New York: Springer.

WALD, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large," *Transactions of the American Mathematical Society*, 54, 426–482.

WATSON, M. W., AND R. F. ENGLE (1985): "Testing for Regression Coefficient Stability with a Stationary AR(1) Process," *Review of Economics and Statistics*, 67, 341–346.

WU, P. X., AND M. L. KING (1994): "One-sided Hypothesis Testing in Econometrics: A Survey," *Pakistan Journal of Statistics*, 10, 261–300.