DEPARTMENT OF ECONOMICS
UNIVERSITY OF CYPRUS

AN ALTERNATIVE ASYMPTOTIC ANALYSIS OF RESIDUAL-BASED STATISTICS

Elena Andreou and Bas J.M. Werker

Discussion Paper 2010-08

P.O. Box 20537, 1678 Nicosia, CYPRUS Tel.: ++357-22893700, Fax: ++357-22895028
Web site: http://www.econ.ucy.ac.cy

An Alternative Asymptotic Analysis of Residual-Based Statistics

Elena Andreou* and Bas J.M. Werker†

University of Cyprus and Tilburg University

This version: June 2010

*Department of Economics, University of Cyprus, P.O.Box 537, CY1678, Nicosia, Cyprus. Tel +357 22 892449, fax +357 22 892432, e-mail Elena.Andreou@ucy.ac.cy.

†Corresponding author: Econometrics and Finance Group, CentER, Tilburg University, P.O.Box 90153, 5000 LE, Tilburg, The Netherlands. Tel +31 13 4662532, fax +31 13 4662875, e-mail Werker@TilburgUniversity.nl.

Abstract

This paper presents an alternative method to derive the limiting distribution of residual-based statistics. Our method does not impose an explicit assumption of (asymptotic) smoothness of the statistic of interest with respect to the model's parameters. and, thus, is especially useful in cases where such smoothness is difficult to establish. Instead, we use a locally uniform convergence in distribution condition, which is automatically satisfied by residual-based specification test statistics. To illustrate, we derive the limiting distribution of a new functional form specification test for discrete choice models, as well as a runs-based tests for conditional symmetry in dynamic volatility models.

JEL codes: C32, C51, C52.

Keywords: Le Cam's third lemma, Local Asymptotic Normality (LAN).

1 Introduction

Residual-based tests are generally used for diagnostic checking of a proposed statistical model. Such specification tests are covered in many textbooks and remain of interest in ongoing research. Similarly, residual-based estimators (often referred to as two-step estimators) are widely applied in econometric work. Traditionally, the asymptotic distribution of residual-based statistics (be it tests or estimators) is derived using a particular model specification, some more or less stringent assumptions about the statistic, and conditions on the first-step estimator employed. A key assumption is some form of (asymptotic) smoothness of the statistic with respect to the parameter to be estimated as formalized first in Pierce (1982) and Randles (1982). Since then, this approach has been significantly extended in, e.g., Pollard (1989), Newey and McFadden (1994), and Andrews (1994).

We present a new and alternative approach that does not involve explicit smoothness conditions for the statistic of interest. Instead, we rely on a locally uniform weak convergence assumption which is shown to be generally (automatically) satisfied by residual-based statistics. Our approach offers a useful and unifying alternative, especially when smoothness conditions are nontrivial to establish or require additional regularity. Some examples of such statistics are, for instance, rank-based statistics (see, e.g., Hallin and Puri, 1991) and statistics based on non-differentiable forecast error loss functions (e.g. McCracker, 2000). Abadie and Imbens (2009) present an application of our method to derive the asymptotic distribution theory of matching estimators based on the estimated propensity score that can be a non-smooth function of the estimated parameters and for which standard bootstrap inference is often not valid (Abadie and Imbens, 2008). In applications where the statistic of interest is smooth, our conditions can be checked along the traditional lines. In order to illustrate our approach, we derive the limiting distribution of a new test based on Kendall's tau for omitted variables in binary choice models and a runs-based test for conditional symmetry in dynamic volatility models.

Our proposed method applies to general model specifications, as long as they satisfy the Uniform Local Asymptotic Normality (ULAN) condition. Most of the standard econometric models satisfy this condition; see Section 3.1 below for a more detailed discussion. The ULAN condition is central in the Hájek and Le Cam's theory of asymptotic statistics (see, e.g., Bickel et al., 1993, Le Cam and Yang, 1990, Pollard, 2004, and van der Vaart, 1998). We use this theory to derive our results. Other advances in econometric theory using the LAN approach can be found in, e.g., Abadir and Distaso (2007), Jeganathan (1995), and Ploberger (2004). For ULAN models, our results offer a simple, yet general, method to derive the asymptotic distribution of residual-based statistics using initial $\sqrt{n}$-consistent estimators. Under the conditions imposed, our main Theorem 3.1 asserts that the residual-based statistic is asymptotically normally distributed with a variance that is a simple function of the limiting variances/covariances of the innovation-based statistic[1], the central sequence (the ULAN equivalent of the derivative of the log-likelihood), and the estimator. Using this approach, we can readily obtain the local power of such residual-based tests, which can also be interpreted in terms of specification tests with locally misspecified alternatives such as in Bera and Yoon (1993). In particular, this allows one to assess in which situations the local power of the residual-based test exceeds, falls below, or equals that of the innovation-based test.

To illustrate our method, we consider two applications. First, we derive the asymptotic distribution of a new nonparametric test for omitted variables in a binary choice model. Second, we discuss a runs-based test for conditional

[1]Throughout the paper, we use the term innovation-based statistic for the statistic applied to the true innovations in the model, i.e., the statistic obtained if the true value of the model parameters were used.

symmetry in dynamic volatility models. These applications purposely focus on non-parametric statistics as these are usually defined in terms of inherently non-smooth statistics like ranks, signs, runs, etc. For these applications, an appropriate form of asymptotic smoothness can probably be established, but our technique offers a useful alternative for which this is not necessary. Our applications are introduced in Section 2. A number of additional applications of our method can be found in Andreou and Werker (2009).

Although the present paper mainly deals with residual-based testing, the results can directly be applied in the area of two-step estimation when assessing the estimation error in a second-step estimator calculated from the residuals of a model estimated in a first step. This problem has received large attention in the econometrics literature, see, e.g, Murphy and Topel (1985, 2002) and Pagan (1986). In the notation below, this would merely mean that the statistic T_n should be taken as the second-step estimation error.

The rest of this paper is organized as follows. The next section introduces the applications we use to illustrate the scope of our technique. Section 3.1 then presents the conditions we need to derive the limiting distribution of a residual-based statistic. Our main result is stated and discussed in Section 3.2. Section 3.3 uses our main theorem to derive the (local) power of residual-based tests and compares this with the local power of the underlying innovation-based tests. We indicate that a technical issue arises when making our ideas rigorous. Section 4 addresses this by discretization and we provide a formal proof of our main result. Section 5 concludes and the appendix contains the proofs and some auxiliary results.

2 Two motivational applications

2.1 Omitted variable test for the Binary Choice model

Consider the binary choice model

$$\mathbb{P}\{Y = 1|X\} = F(X^T\theta), \tag{2.1}$$

where Y denotes a binary response variable, X some exogenous explanatory variables, and F a given probability distribution function. We assume that the distribution function F admits a continuous density f and that the Fisher Information matrix

$$I_F(\theta) = \mathbb{E}\frac{f(X^T\theta)^2}{F(X^T\theta)\,(1 - F(X^T\theta))}XX^T, \tag{2.2}$$

exists and is continuous in θ. For inference, an i.i.d. sample of observations (Y_i, X_i), $i = 1, \ldots, n$, is available.

The generalized residuals, for given parameter value θ, are defined as

$$\varepsilon_i^G(\theta) = \frac{Y_i - F(X_i^T\theta)}{F(X_i^T\theta)\,(1 - F(X_i^T\theta))} f(X_i^T\theta) \tag{2.3}$$

The classical test for functional specification checks for a possibly omitted variable Z_i using the statistic

$$T_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i^G(\theta) Z_i. \tag{2.4}$$

The statistic $T_n(\theta)$ is innovation based as it depends on the unknown true value of the parameter θ. The limiting distribution of this innovation-based statistic follows immediately from the classical Central Limit Theorem as soon as $\varepsilon_i^G(\theta)Z_i$ has finite variance and zero mean.

In applications, the unknown parameter θ is replaced by an estimator $\hat{\theta}_n$, for instance, the maximum likelihood estimator $\hat{\theta}_n^{(ML)}$. This leads to the residual-based statistic $T_n(\hat{\theta}_n)$. The traditional way of deriving the limiting distribution of $T_n(\hat{\theta}_n)$ relies on linearizing the statistic $T_n(\theta)$; see, for instance, Pagan and Vella (1989). This approach leads to

$$T_n(\hat{\theta}_n) \xrightarrow{\mathcal{L}} N\left(0, \mathbb{E}WZZ^T - \mathbb{E}WXZ^T \left(\mathbb{E}WXX^T\right)^{-1} \mathbb{E}WXZ^T\right), \tag{2.5}$$

as $n \to \infty$, with

$$W = \frac{f(X^T\theta)^2}{F(X^T\theta)\,(1 - F(X^T\theta))}.$$

The test statistic (2.4) checks for linear correlation between the generalized residuals and the possibly omitted variable Z. One could also be interested in a test with power against nonlinear forms of dependence based on Kendall's tau applied to the pairs $(\varepsilon_i^G(\theta), Z_i)$. For simplicity we consider the case where the possibly omitted variables are univariate, i.e., $Z_i \in \mathbb{R}$. Recall that the population version of Kendall's tau is defined as

$$\tau = 4\mathbb{P}\{\varepsilon_i^G(\theta) < \varepsilon_j^G(\theta),\ Z_i < Z_j\} - 1,\ i \neq j. \tag{2.6}$$

An appropriately scaled innovation-based version of Kendall's tau is the U-statistic

$$\begin{aligned} T_n^\tau(\theta) &= \sqrt{n}\begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{i=1}^{n}\sum_{j=1}^{i-1}\left[4I\{\varepsilon_i^G(\theta) < \varepsilon_j^G(\theta),\ Z_i < Z_j\} - 1\right] \\ &\xrightarrow{\mathcal{L}} N\left(0, \frac{4}{9}\right). \end{aligned}$$

This limiting distribution, under the null hypothesis of independent ε_i^G and Z_i, can be obtained using the projection theorem for U-statistics, e.g., Theorem 12.3 in van der Vaart (1998).

Deriving the limiting distribution of the residual-based statistic $T_n^\tau(\hat{\theta}_n)$ using linearizion is less obvious due to the inherent non-differentiability of the indicator functions in $T_n^\tau(\theta)$. Our approach to residual-based statistics will give this limiting distribution at about the same effort as the smooth classical statistic $T_n(\theta)$. More precisely, using our technique we show, in Section 3,

$$T_n^\tau(\hat{\theta}_N^{(ML)}) \xrightarrow{\mathcal{L}} N\left(0; \frac{4}{9} - \alpha^T I_F(\theta)\alpha\right), \tag{2.7}$$

with α defined in (3.5). This is not only a useful result that shows how the asymptotic distribution of Kendall's tau test statistic differs when applied to residuals (instead of innovations), but also a practical result given that the asymptotic variance in (2.7) can easily be estimated consistently (see Section 3 for details). This test complements existing tests in the literature. □

2.2 Runs test for symmetry in Dynamic Volatility models

Consider the following time series model

$$Y_t = \sigma_{t-1}(\theta)\varepsilon_t,\ t = 1, \ldots, n, \tag{2.8}$$

where $\sigma_{t-1}(\theta)$ depends on past values $Y_{t-1}, Y_{t-2}, \ldots$ and $\{\varepsilon_t\}$ is a sequence of i.i.d. innovations. Assume that these innovations ε_t have an absolutely continuous density f with finite Fisher information for location and scale, i.e., $I_l := \int (f'(x)/f(x))^2 f(x)\mathrm{d}x < \infty$ and $I_s := \int (1 + xf'(x)/f(x))^2 f(x)\mathrm{d}x <$

∞. Finally, impose $\mathbb{E}\varepsilon_t = 0$, $\mathbb{E}\varepsilon_t^2 = 1$, and $\kappa_\varepsilon := \mathbb{E}\varepsilon_t^4 < \infty$ and assume that a stationary and ergodic solution to (2.8) exists. These are standard assumptions for most GARCH-type models.

Many specification tests concerning the innovations in stochastic volatility models have been introduced and studied in the literature. We consider a nonparametric test of conditional symmetry based on Wald-Wolfowitz runs. One advantage of such a test is that it does not require the existence of any higher order moments of the innovation distribution and, thus, can be considered robust to different distributions and outliers. This is particularly relevant given that there is no consensus in the empirical literature as to the form of heavy-tailed distributions in, e.g., financial time series. This test for conditional symmetry counts the number of runs of all negative or all positive residuals. Formally, defining $I_t(\theta) = I\{\varepsilon_t(\theta) < 0\}$, the test statistic becomes

$$T_n(\theta) = \frac{1}{\sqrt{n}} \sum_{t=2}^{n} \left([I_t(\theta) - I_{t-1}(\theta)]^2 - \frac{1}{2} \right). \tag{2.9}$$

This is a simple nonparametric test that complements existing tests for symmetry such as, for instance, in Bai and Ng (2001) and Bera and Premarantne (2005, 2009).

Using standard central limit results, one easily finds that the limiting null distribution of the innovation-based statistic $T_n(\theta)$ is $N(0, 1/4)$. Our detailed results in Section 3 show that this asymptotic null distribution needs not be adapted when applied to residuals of dynamic volatility models. The results in Section 3.3 furthermore show that the asymptotic local power of this runs test is the same whether applied to innovations or to residuals. □

3 Main results

Our results are derived using the Hájek and Le Cam techniques of asymptotic statistics. We first introduce the assumptions needed. Subsequently, we will derive the asymptotic size and local power of residual-based tests.

3.1 Assumptions

Let us formally introduce the models we consider in this paper. Thereto, let $\mathcal{E}^{(n)}$ denote a sequence of experiments defined on a common parameter set $\Theta \subset \mathbb{R}^k$:

$$\mathcal{E}^{(n)} = \left\{ X^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)} = \left(\mathbb{P}_\theta^{(n)} \ : \ \theta \in \Theta \right) \right\},$$

where $\left(X^{(n)}, \mathcal{A}^{(n)}\right)$ is a sequence of measurable spaces and, for each n and $\theta \in \Theta$, $\mathbb{P}_\theta^{(n)}$ a probability measure on $\left(X^{(n)}, \mathcal{A}^{(n)}\right)$. We assume throughout this paper that Θ is a subset of $\mathbb{R}^k$ so that we consider the effect of pre-estimating a Euclidean parameter. We also assume that pertinent asymptotics in the sequence of experiments take place at the usual $\sqrt{n}$ rate. Other rates, as occur, for instance, in non-stationary time series, can be easily adopted at the cost of more cumbersome notation only.

Our analysis is based on two assumptions. The first is Condition (ULAN) that imposes regularity on the model at hand. This condition involves neither the statistic T_n nor the estimator $\hat{\theta}_n$. During the last thirty years, the ULAN condition has been established for most standard cross-section and time-series models. To introduce the condition, let $\theta_0 \in \Theta$ denote a fixed value of the Euclidean parameter and let (θ_n) and (θ'_n) denote sequences contiguous to θ_0, i.e., $\delta_n = \sqrt{n}(\theta_n - \theta_0)$ and $\delta'_n = \sqrt{n}(\theta'_n - \theta_0)$ are bounded in $\mathbb{R}^k$. Write $\Lambda^{(n)}(\theta'_n|\theta_n) = \log\left(\mathrm{d}\mathbb{P}^{(n)}_{\theta'_n}/\mathrm{d}\mathbb{P}^{(n)}_{\theta_n}\right)$ for the log-likelihood of $\mathbb{P}^{(n)}_{\theta'_n}$ with respect to $\mathbb{P}^{(n)}_{\theta_n}$. In case $\mathbb{P}^{(n)}_{\theta'_n}$ is not dominated by $\mathbb{P}^{(n)}_{\theta_n}$, we mean the Radon-Nikodym derivative of the absolute continuous part in the Lebesgue decomposition of $\mathbb{P}^{(n)}_{\theta'_n}$ with respect to $\mathbb{P}^{(n)}_{\theta_n}$ (see Strasser, 1985, Definition 1.3).

Condition (ULAN): The sequence of experiments $\mathcal{E}^{(n)}$ is Uniformly Locally Asymptotically Normal (ULAN) in the sense that there exists a sequence of random variables $\Delta^{(n)}(\theta)$ (the *central sequence*) such that for all sequences θ_n and θ'_n contiguous to θ_0, we have

$$\begin{aligned}
\Lambda^{(n)}(\theta'_n|\theta_n) &= (\delta'_n - \delta_n)^T \Delta^{(n)}(\theta_0 + \delta_n/\sqrt{n}) - \frac{1}{2}(\delta'_n - \delta_n)^T I_F (\delta'_n - \delta_n) + o_{\mathrm{P}}(1) \\
&= (\delta'_n - \delta_n)^T \Delta^{(n)}(\theta_0 + \delta'_n/\sqrt{n}) + \frac{1}{2}(\delta'_n - \delta_n)^T I_F (\delta'_n - \delta_n) + o_{\mathrm{P}}(1),
\end{aligned}$$

where $\theta_n = \theta_0 + \delta_n/\sqrt{n}$, $\theta'_n = \theta_0 + \delta'_n/\sqrt{n}$, and both δ_n and δ'_n are bounded sequences. The central sequence $\Delta^{(n)}(\theta_n)$ is asymptotically normally distributed with zero mean and variance $I_F(\theta_0)$, i.e., $\Delta^{(n)}(\theta_n) \xrightarrow{\mathcal{L}} N(0, I_F)$, under $\mathbb{P}^{(n)}_{\theta_n}$, as $n \to \infty$. Here, $I_F(\theta_0)$ is called the Fisher information matrix (at $\theta = \theta_0$). $\square$

Remark 3.1. The central sequence $\Delta^{(n)}(\theta_n)$ is the ULAN equivalent of the derivative of the log-likelihood function. The formulation in Condition (ULAN) allows for situations where the log-likelihood is not point-wise differentiable, e.g., when using double-exponential densities. The 'uniformity' in the ULAN condition lies in the use of contiguous alternatives θ_n in the denominator of the log-likelihood $\Lambda^{(n)}(\theta'_n|\theta_n)$. In case the likelihood expansion is only required for $\theta_n = \theta_0$, the condition is called Local Asymptotic Normality (LAN). The ULAN condition can usually be established under the same conditions as

LAN. We need the uniform version in the proof of our main result essentially due to the fact that residual-based statistics are calculated using residuals based on (random) local alternatives $\hat{\theta}_n$. Note that the ULAN condition as such is simply an assumption relating to the model and does not affect the (non)smoothness assumption of the statistic one wishes to consider. $\square$

Remark 3.2. The ULAN condition presents a prime example in the theory of convergence of statistical experiments. The quadratic expansion of the log-likelihood ratio in the local parameter $\delta_n' - \delta_n$ is equal to the log-likelihood ratio in the Gaussian shift model $\{N(I_F(\theta_0)^{-1}\delta, I_F(\theta_0)^{-1}) : \ \delta \in \mathbb{R}^k\}$. This can be shown to imply that the sequence of localized experiments $\{\mathbb{P}^{(n)}_{\theta_n+\delta} : \ \delta \in \mathbb{R}^k\}$ converges, in an appropriate sense, to the Gaussian shift experiment. This in turn implies that asymptotic analysis in the original experiments can be based on properties of the limiting Gaussian shift model. A useful consequence of the ULAN, or even LAN, property is that the sequences $\mathbb{P}^{(n)}_{\theta_n'}$ and $\mathbb{P}^{(n)}_{\theta_n'}$ are contiguous (see, e.g., Le Cam and Yang, 1990, or van der Vaart, 1998). As a result, convergence in probability under $\mathbb{P}^{(n)}_{\theta_n}$ is equivalent to convergence under $\mathbb{P}^{(n)}_{\theta_n'}$. In particular, any $o_{\mathrm{P}}(1)$-terms in Condition (ULAN), and in the remainder of this paper, hold simultaneously under $\mathbb{P}^{(n)}_{\theta_0}$, $\mathbb{P}^{(n)}_{\theta_n}$, and $\mathbb{P}^{(n)}_{\theta_n'}$. $\square$

Both the binary choice and the dynamic volatility model introduced in Section 2 satisfy the ULAN condition; this is discussed below.

Condition (ULAN) for the Binary Choice model: The log-likelihood function for the Binary Choice model (2.1) is given by

$$\log L(\theta) = \sum_{i=1}^{n} Y_i \log F(X_i^T\theta) + (1 - Y_i)\log\left(1 - F(X_i^T\theta)\right). \tag{3.1}$$

Under the assumption imposed on the Fisher information matrix (2.2), Condition (ULAN) is easily seen to be satisfied as Proposition 2.1.1 in Bickel et al. (1993) applies. This proposition establishes ULAN for models with i.i.d. observations using the so-called Differentiability in Quadratic Mean (DQM) condition that is obviously satisfied in the binary choice model. The central sequence is given by

$$\Delta^{(n)}(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{Y_i - F(X_i^T\theta)}{F(X_i^T\theta)\left(1 - F(X_i^T\theta)\right)} f(X_i^T\theta)X_i = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \varepsilon_i^G(\theta)X_i. \tag{3.2}$$

The corresponding Fisher information matrix, at θ, is indeed $I_F(\theta)$ as follows from the central limit theorem. □

Condition (ULAN) for the Dynamic Volatility model: For various specifications of the conditional volatility $\sigma_{t-1}(\theta)$, the induced parametric model satisfies Condition (ULAN). In particular this holds for the classic GARCH(1,1) model where $\sigma_t^2(\theta) = \omega + \alpha Y_{t-1}^2 + \beta\sigma_{t-1}^2(\theta)$ with $\theta = (\omega, \alpha, \beta) \in \mathbb{R}_+^3$ under the Nelson (1990) condition for strict stationarity:$\mathbb{E}\log(\beta + \alpha\varepsilon_t^2) < 0$ (Theorem 2.1in Drost and Klaassen, 1997). Note that IGARCH(1,1) models, for which $\alpha + \beta = 1$, are not ruled out. Moreover, Condition (ULAN) holds for (G)ARCH-in-mean models (Linton, 1993, and Drost and Klaassen, 1997) and the Asymmetric GARCH model (Sun and Stengos, 2006). In all cases, the central sequence for θ is given by

$$\Delta^{(n)}(\theta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} -\frac{1}{2}\left(1 + \varepsilon_t(\theta)\frac{f'(\varepsilon_t(\theta))}{f(\varepsilon_t(\theta))}\right)\frac{\partial}{\partial\theta}\log\sigma_{t-1}^2(\theta),$$

where $\varepsilon_t(\theta) := Y_t/\sigma_{t-1}(\theta)$. The Fisher information is given by $I_F = I_s A(\theta)$ with

$$A(\theta) = \lim_{n\to\infty} \frac{1}{4n}\sum_{t=1}^{n} \frac{\partial}{\partial\theta}\log\sigma_{t-1}^2(\theta)\frac{\partial}{\partial\theta^T}\log\sigma_{t-1}^2(\theta),$$

where the limit is taken in probability. □

Our second assumption is about the statistic of interest $T_n(\theta)$ and the (first-step) estimator $\hat{\theta}_n$ for θ used. Formally, we are interested in some innovation-based statistic $T_n(\theta)$, depending on the unknown model parameter θ. Generally, the asymptotic behavior of this innovation-based statistic follows easily from classical limit arguments. The focus of interest of the present paper is the asymptotic behavior of the residual-based statistic $T_n(\hat{\theta})$ obtained by replacing the true value of θ by some estimate $\hat{\theta}_n$.

Traditionally, following Pierce (1982) and Randles (1982), several papers analyze the asymptotic behavior of the residual-based statistic relying on a condition like

$$T_n(\theta_0 + \delta_n/\sqrt{n}) = T_n(\theta_0) - c^T\delta_n + o_{\mathrm{P}}(1), \tag{3.3}$$

under $\mathbb{P}^{(n)}_{\theta_0}$, for bounded sequences δ_n. Condition (3.3) is sometimes reinforced to hold for random sequences $\hat{\delta}_n = O_{\mathrm{P}}(1)$. However, this reinforcement is generally not required if one would resort to discretized estimators as in Section 4. In the ubiquitous case that the statistic $T_n(\theta)$ is, up to $o_p(1)$-terms, some average $n^{-1/2}\sum_{t=1}^n m(X_t;\theta)$ of expectation-zero functions m of observations $X_1,\dots,X_n$, Condition (3.3) is verified if m is differentiable in θ (a route, for instance, followed in Newey and McFadden, 1994) or if $\mathbb{E}m(X_t;\theta)$ is differentiable in θ as in, e.g., the empirical process approach of Andrews (1994).

Given (3.3), the limiting distribution of $T_n(\hat{\theta}_n)$ is subsequently obtained, under θ_0, from

$$
\begin{aligned}
T_n\left(\hat{\theta}_n\right) &= T_n\left(\theta_0 + \sqrt{n}\left(\hat{\theta}_n - \theta_0\right)/\sqrt{n}\right) \\
&= T_n(\theta_0) - c^T\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) + o_{\mathrm{P}}(1). \qquad (3.4)
\end{aligned}
$$

Given this expansion, the asymptotic distribution of $T_n(\hat{\theta}_n)$ follows immediately from the joint limiting distribution of the innovation-based statistic $T_n(\theta_0)$ and the estimation error $\sqrt{n}(\hat{\theta}_n - \theta_0)$, combined with the knowledge of c. Instead of the smoothness (3.3), we impose joint Asymptotic Normality (AN) on the estimator $\hat{\theta}_n$, the innovation-based statistic $T_n(\theta)$, and the central sequence $\Delta^{(n)}(\theta)$.

Condition (AN): Consider a sequence θ_n contiguous to θ_0. The innovation-based test statistic $T_n(\theta_n)$, the central sequence $\Delta^{(n)}(\theta_n)$, and the estimation error $\sqrt{n}(\hat{\theta}_n - \theta_n)$ are jointly asymptotically normally distributed, under $\mathbb{P}^{(n)}_{\theta_n}$, as $n \to \infty$ and as $\delta_n \to \delta$. More precisely,

$$\begin{bmatrix} T_n(\theta_n) \\ \\ \Lambda^{(n)}(\theta_n|\theta_0) \\ \\ \sqrt{n}(\hat{\theta}_n - \theta_n) \end{bmatrix} \xrightarrow{\mathcal{L}} \begin{bmatrix} T \\ \\ \frac{1}{2}\delta^T I_F \delta + \delta^T \Delta \\ \\ Z \end{bmatrix}$$

$$\sim \quad N\left(\begin{bmatrix} 0 \\ \\ \frac{1}{2}\delta^T I_F \delta \\ \\ 0 \end{bmatrix} ; \begin{bmatrix} \tau^2 & c^T\delta & \alpha^T \\ \\ \delta^T c & \delta^T I_F \delta & \delta^T \\ \\ \alpha & \delta & \Gamma \end{bmatrix} \right).$$

□

Remark 3.3. Observe that the use of the notation c in the derivative in (3.3) is consistent with the use of c in Condition (AN). This can be seen as follows. As τ^2 denotes the limiting variance of $T_n(\theta_0)$ under $\mathbb{P}^{(n)}_{\theta_0}$, the limiting distribution of $T_n(\theta_n)$ with $\theta_n = \theta_0 + \delta/\sqrt{n}$, under $\mathbb{P}^{(n)}_{\theta_0}$, follows from (3.3) as $N(-c^T\delta, \tau^2)$. However, as in the proof of Theorem 3.1 below, it also follows from Le Cam's third lemma, as recalled in Appendix A, as $N(-c^T\delta, \tau^2)$. □

Condition (AN) requires a locally uniform version of the central limit theorem as we consider convergence, under $\mathbb{P}^{(n)}_{\theta_n}$, of statistics evaluated at $\theta = \theta_n$. Such a condition is clearly stronger than convergence at and under a single fixed θ_0. However, Appendix B gives two results that effectively reduce the technical burden to an analysis of $T_n(\theta_0)$ and $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$ under $\mathbb{P}^{(n)}_{\theta_0}$ only. These results are useful in both applications we consider.

Condition (AN) for the Binary Choice model: We consider the situation where θ is estimated using maximum likelihood, so that $\sqrt{n}(\hat{\theta}_n - \theta_0) = I_F(\theta_0)^{-1}\Delta^{(n)}(\theta_0) + o_{\mathrm{P}}(1)$. Using Proposition B.2, Condition (AN) follows from Condition (ULAN) as far as $\Lambda^{(n)}(\theta_n|\theta_0)$ and $\sqrt{n}(\hat{\theta}_n - \theta_n)$ are concerned. Clearly, we have $\Gamma = I_F^{-1}$ and $\alpha = \Gamma c$. Concerning the innovation-based statistic Proposition B.1 applies.Using maximum likelihood implies $\Gamma = I_F^{-1}$ and $\Gamma c = \alpha$. Finally, α itself follows from

$$
\begin{aligned}
\alpha &= \lim_{n\to\infty} \mathrm{Cov}_{\theta_0}\left\{\sqrt{n}(\hat{\theta}_n - \theta_0), T_n^\tau(\theta_0)\right\} \qquad (3.5) \\
&= I_F^{-1} \lim_{n\to\infty} \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{l=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{i-1} \mathrm{Cov}\left\{\varepsilon_l^G X_l, 4I\{\varepsilon_i^G(\theta) < \varepsilon_j^G(\theta),\ Z_i < Z_j\} - 1\right\} \\
&= 4I_F^{-1} \lim_{n\to\infty} \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{i=1}^{n}\sum_{j=1}^{i-1} \mathbb{E}\left(\varepsilon_i^G X_i + \varepsilon_j^G X_j\right) I\{\varepsilon_i^G(\theta) < \varepsilon_j^G(\theta),\ Z_i < Z_j\}.
\end{aligned}
$$

Further simplification of this expression is not necessary as it is easily estimated consistently. □

Condition (AN) for the Dynamic Volatility model: For conditionally heteroskedastic models, the QMLE estimator $\widehat{\theta}_n$ based on an assumed Gaussian distribution for the innovations ε_t is a popular choice. For the GARCH(1,1) model, Lumsdaine (1996) establishes consistency and asymptotic normality of this estimator under conditions implying the ones imposed above. Her results essentially also establish (3.6) below. Recently, Berkes and Horváth (2004) have improved upon these results showing, for GARCH(p,q) processes, an asymptotically linear representation (B.4) for given θ:

$$\sqrt{n}(\widehat{\theta}_n - \theta) = -A(\theta)^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{1}{4} \left(1 - \frac{Y_t^2}{\sigma_{t-1}^2(\theta)}\right) \frac{\partial}{\partial \theta} \log \sigma_{t-1}^2(\theta) + o_{\mathrm{P}}(1), \quad (3.6)$$

under $\mathbb{P}_\theta^{(n)}$. More precisely, (3.6) follows from their result (4.18) which, as noted in the proof of their Theorem 2.1 is also valid for $\widehat{\theta}_n$, applied to their Example 2.1. From the representation (3.6) one finds the asymptotic variance of the QMLE estimator as

$$\Gamma = \Gamma(\theta) = \frac{\kappa_\varepsilon - 1}{4} A(\theta)^{-1}, \quad (3.7)$$

with $\kappa_\varepsilon = \mathbb{E}\varepsilon_t^4$ (compared to Theorem 1.2 in Berkes and Horváth, 2004).

These results can be reinforced to obtain an estimator that actually satisfies the local uniformity in Condition (AN) if (B.5) holds. While neither Lumsdaine (1996) nor Berkes and Horváth (2004) explicitly mention (B.5), their results allow us to invoke Proposition B.2 as the following lemma shows.

Lemma 3.1. *For the GARCH(p, q) model, (B.5) holds for the Gaussian QMLE estimator with*

$$\psi_t(\theta) = -A(\theta)^{-1} \frac{1 - \varepsilon_t^2(\theta)}{4} \frac{\partial}{\partial \theta} \log \sigma_{t-1}^2(\theta).$$

Proof. First of all note that $A(\theta)$ is invertible and continuous by applying the mean-value theorem and using Lemma 3.6 in Berkes and Horváth (2004). As a result, it suffices to study, for local alternatives θ_n to θ_0,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \frac{1 - \varepsilon_{t-1}^2(\theta_n)}{4} \frac{\partial}{\partial \theta} \log \sigma_{t-1}^2(\theta_n) - \frac{1 - \varepsilon_{t-1}^2(\theta_0)}{4} \frac{\partial}{\partial \theta} \log \sigma_{t-1}^2(\theta_0) \right\}, \tag{3.8}$$

under $\mathbb{P}_{\theta_0}^{(n)}$. Consider a given element $\theta^{(j)}$ of the vector θ. Applying Taylor's theorem to (3.8) we find, for the element corresponding to $\theta^{(j)}$,

$$\left(\frac{1}{n} \sum_{t=1}^{n} \frac{\varepsilon_t^2(\theta_n')}{4} \frac{\partial}{\partial \theta} \log \sigma_{t-1}^2(\theta_n') \frac{\partial}{\partial \theta^T} \log \sigma_{t-1}^2(\theta_n') + \frac{1 - \varepsilon_t^2(\theta_n')}{4} \frac{\partial^2}{\partial \theta \partial \theta^T} \log \sigma_{t-1}^2(\theta_n') \right) \\ \times \sqrt{n}(\theta_n - \theta_0),$$

with θ_n' on the line segment from θ_0 to θ_n. Given the boundedness of $\sqrt{n}\theta_n - \theta_0)$, it suffices to show that the term in parentheses converges to $A(\theta_0)$. This, however, follows from Lemma 4.4 in Berkes and Horváth (2004). $\square$

This shows that, with respect to the initial estimator $\widehat{\theta}_n$, Condition (AN) is indeed satisfied. With respect to the log-likelihood, as mentioned before, Condition (AN) follows immediately from Condition (ULAN) and with respect to the runs statistic, Proposition B.1 applies once more. □

3.2 Size of residual-based tests

We now state and prove, in an informal way, the main result of the paper. All statements will be made precise in Section 4.

Theorem 3.1. *Under the Conditions (ULAN) and (AN) and in a way that will be made precise in Section 4, we have, for the residual-based statistic* $T_n(\hat{\theta}_n)$*,*

$$\begin{aligned} T_n(\hat{\theta}_n) &\xrightarrow{\mathcal{L}} N\left(0, \tau^2 + (\alpha - \Gamma c)^T \Gamma^{-1} (\alpha - \Gamma c) - \alpha^T \Gamma^{-1} \alpha\right) \quad (3.9) \\ &= N\left(0, \tau^2 + c^T \Gamma c - 2\alpha^T c\right), \end{aligned}$$

under $\mathbb{P}_{\theta_n}^{(n)}$*, as* $n \to \infty$*.*

Proof. (Intuition) Introduce the distribution

$$\begin{bmatrix} T \\ \Delta \\ Z \end{bmatrix} \sim N\left(0, \begin{bmatrix} \tau^2 & c^T & \alpha^T \\ c & I_F & I_k \\ \alpha & I_k & \Gamma \end{bmatrix}\right). \quad (3.10)$$

where I_k denotes the $k \times k$ identity matrix. From Condition (AN), applied with θ_n replaced by $\theta_n + \delta/\sqrt{n}$ and using Condition (ULAN), we have for all $\delta \in \mathbb{R}^k$, under $\mathbb{P}^{(n)}_{\theta_n+\delta/\sqrt{n}}$ and as $n \to \infty$,

$$\begin{bmatrix} T_n(\theta_n + \delta/\sqrt{n}) \\ \Lambda^{(n)}(\theta_n|\theta_n + \delta/\sqrt{n}) \\ \sqrt{n}(\hat{\theta}_n - \theta_n - \delta/\sqrt{n}) \end{bmatrix} \xrightarrow{\mathcal{L}} \begin{bmatrix} T \\ -\frac{1}{2}\delta^T I_F \delta - \delta^T \Delta \\ Z \end{bmatrix},$$

while, as a consequence of Le Cam's third lemma, the same vector converges under $\mathbb{P}^{(n)}_{\theta_n}$ to

$$\begin{bmatrix} T - c^T\delta \\ +\frac{1}{2}\delta^T I_F \delta - \delta^T \Delta \\ Z - \delta \end{bmatrix}.$$

The quantity of interest can now be written, for $t \in \mathbb{R}$,

$$\begin{aligned}
&\mathbb{P}^{(n)}_{\theta_n}\left\{T_n(\hat{\theta}_n) \le t\right\} \\
&= \int_{\delta\in\mathbb{R}^k} \mathbb{P}^{(n)}_{\theta_n}\left\{T_n(\hat{\theta}_n) \le t | \hat{\theta}_n = \theta_n + \delta/\sqrt{n}\right\} \mathrm{d}\mathbb{P}^{(n)}_{\theta_n}\left\{\hat{\theta}_n \le \theta_n + \delta/\sqrt{n}\right\} \\
&= \int_{\delta\in\mathbb{R}^k} \mathbb{P}^{(n)}_{\theta_n}\left\{T_n(\theta_n + \delta/\sqrt{n}) \le t | \sqrt{n}(\hat{\theta}_n - \theta_n) = \delta\right\} \mathrm{d}\mathbb{P}^{(n)}_{\theta_n}\left\{\sqrt{n}(\hat{\theta}_n - \theta_n) \le \delta\right\} \\
&\to \int_{\delta\in\mathbb{R}^k} \mathbb{P}\left\{T - c^T\delta \le t | Z = \delta\right\} \mathrm{d}\mathbb{P}\left\{Z \le \delta\right\} \\
&= \int_{\delta\in\mathbb{R}^k} \Phi\left(\frac{t + (c - \Gamma^{-1}\alpha)^T\delta}{\sqrt{\tau^2 - \alpha^T\Gamma^{-1}\alpha}}\right) \mathrm{d}\mathbb{P}\left\{Z \le \delta\right\},
\end{aligned}$$

where Φ denotes the cumulative distribution function of the standard normal distribution and we used the result that, conditionally on $Z = z$, T has a $N(\alpha^T\Gamma^{-1}z, \tau^2 - \alpha^T\Gamma^{-1}\alpha)$ distribution. Observe that, if we introduce the distribution

$$\begin{bmatrix} X \\ Z \end{bmatrix} \sim N\left(0, \begin{bmatrix} \tau^2 - \alpha^T\Gamma^{-1}\alpha + (\alpha - \Gamma c)^T\Gamma^{-1}(\alpha - \Gamma c) & (\alpha - \Gamma c)^T \\ (\alpha - \Gamma c) & \Gamma \end{bmatrix}\right),$$

the distribution of X conditionally on $Z = \delta$ is $N(-(c - \Gamma^{-1}\alpha)^T\delta, \tau^2 - \alpha^T\Gamma^{-1}\alpha)$. Consequently, the limit of $\mathbb{P}^{(n)}_{\theta_n}\left\{T_n(\hat{\theta}_n) \leq t\right\}$ can be written as

$$\int_{\delta\in\mathbb{R}^k} \mathbb{P}\left\{X \leq t | Z = \delta\right\} \mathrm{d}\mathbb{P}\left\{Z \leq \delta\right\} = \mathbb{P}\left\{X \leq t\right\},$$

from which (3.9) follows. □

Remark 3.4. Note that the limiting distribution (3.9) of the residual-based statistic does not depend on the actual sequence θ_n, but only on its limit θ_0 (through the covariances τ^2, c, α, and Γ). In particular this limiting distribution is also valid for $\theta_n = \theta_0$. Local parameter changes *within* the model specified thus, indeed, do not affect the limiting distribution of the residual-based statistic. The local power of $T_n(\hat{\theta}_n)$ for alternatives *outside* the model specified is studied in Section 3.3. □

Remark 3.5. In the above informal proof the convergence of the conditional distribution $\mathbb{P}^{(n)}_{\theta_n}\left\{T_n(\theta_n + \delta/\sqrt{n}) \leq t | \sqrt{n}(\hat{\theta}_n - \theta_n) = \delta\right\}$ to the limit $\mathbb{P}\left\{T - c^T\delta \leq t | Z = \delta\right\}$ is the most delicate part, since the convergence takes place in the conditioning event as well. A formalization of such a convergence would require conditions under which a conditional probability or expectation is continuous with respect to the conditioning event. This question has been studied in the literature by introducing various topologies on the space of conditioning σ-fields. A good reference is the paper by Cotter (1986) that compares some topologies. From our point of interest, Cotter (1986) essentially shows that the required continuity property only holds for discrete probability distributions. Indeed, we formalize Theorem 3.1 by discretizing the estimator $\hat{\theta}_n$ appropriately. See Section 4 for details. □

Remark 3.6. Theorem 3.1 has been stated for univariate statistics $T_n(\theta)$, but can easily be extended to the multivariate case using the Cramér-Wold device. In such a multivariate setting τ^2, c, and α in Condition (AN) are matrices. By taking arbitrary linear combinations of the components of T_n and applying the univariate version of Theorem 3.1, we find that the same limiting distribution (3.9) holds with τ^2 replaced by the limiting variance matrix of T_n, c the limiting covariance matrix between the statistic and the central sequence, and α the limiting covariance matrix between the statistic and the estimator used. □

Size of Kendall's tau for the Binary Choice model: Recall that the limiting variance of the innovation-based Kendall's tau is $\tau^2 = 4/9$. In order to derive the appropriate variance correction when calculating Kendall's tau using generalized residuals (calculated on the basis of the maximum likelihood estimator $\hat{\theta}_n^{(ML)}$), we adopt Theorem 3.1. As $\alpha = \Gamma c$, applying Theorem 3.1 we obtain, under the null hypothesis of no omitted variables,

$$T_n^\tau(\hat{\theta}_n^{(ML)}) \xrightarrow{\mathcal{L}} N\left(0; \frac{4}{9} - \alpha^T I_F \alpha\right).$$

This nonparametric test for omitted variables in the binary choice model has not been considered before in the literature. We provide this application to show that its limiting distribution is easily derived in the framework we propagate. □

Size of runs test for the Dynamic Volatility model: Without going through all the calculations in detail, observe that in this case $c = 0$ since

$$\begin{aligned}
\mathbb{E}I\{\varepsilon_t < 0\}\left(1 + \varepsilon_t f'(\varepsilon_t)/f(\varepsilon_t)\right) &= \frac{1}{2} + \int_{x=-\infty}^{0} x f'(x)\mathrm{d}x \\
&= \frac{1}{2} - \int_{x=-\infty}^{0} f(x)\mathrm{d}x = 0,
\end{aligned}$$

which implies

$$\mathbb{E}_{t-1}[I\{\varepsilon_t < 0\} - I\{\varepsilon_{t-1} < 0\}]^2\left(1 + \varepsilon_t f'(\varepsilon_t)/f(\varepsilon_t)\right) = 0.$$

Consequently, using Theorem 3.1, the asymptotic covariance of the Wald-Wolfowitz runs test for symmetry need not be adapted to estimation error when applied to the residuals of dynamic volatility models. □

If we think of canonical applications, $T_n(\theta)$ represents a test statistic for distributional or time series properties of some innovations in the model, while $T_n(\hat{\theta})$ denotes the same statistic applied to estimated residuals in the model. Theorem 3.1 shows that replacing innovations by residuals may leave the asymptotic variance of the test statistic unchanged, increase it, or decrease it, depending on the value of $(\alpha - \Gamma c)^T \Gamma^{-1} (\alpha - \Gamma c)$ as compared to $\alpha^T \Gamma^{-1} \alpha$. Several special cases can occur.

First, if $c = 0$, the residual-based statistic has the same asymptotic variance as the statistic based on the true innovations. In particular, no adaptation is necessary in critical values in order to guarantee the appropriate asymptotic size of the test when applied to estimated residuals. However, the power of the residual-based test $T_n(\hat{\theta}_n)$ may be different from that of the innovation-based test $T_n(\theta_n)$ (see Section 3.3 for details). Recall that, under $c = 0$, the test statistic and the central sequence are asymptotically independent. As a result, the distribution of the test statistic $T_n(\theta_0)$ is insensitive to local changes in the parameter θ. In particular, the asymptotic distribution of $T_n(\theta_0)$ is the same under all probability distributions $\mathbb{P}_{\theta_n}^{(n)}$, whatever the local parameter sequence θ_n. Or, equivalently in our setup, the asymptotic distribution of $T_n(\theta_n)$ under $\mathbb{P}_{\theta_0}^{(n)}$ is the same for each local sequence θ_n. As estimated parameter values $\hat{\theta}$ differ from θ_0 in the order of magnitude of $1/\sqrt{n}$, these remarks apparently carry over to the residual-based statistic. Our runs-based test for symmetry in the dynamic volatility model falls under this scheme.

A second special case occurs if $\alpha = \Gamma c$. This happens, for instance, when the initial estimator $\hat{\theta}_n$ is efficient (as in that case $\Gamma = I_F^{-1}$ and $\alpha = I_F^{-1} c$). In such situation, the limiting variance of the residual-based statistic is smaller than the limiting variance of the statistic applied to the true innovations, with strict inequality if $\alpha \neq 0$. Pierce (1982) restricts attention to this efficient initial estimator case and, imposing a differentiability condition on $T_n(\theta)$, finds the same reduction in the limiting variance. This occurs, for instance, in our application of omitted variables tests in the binary choice model.

Finally, it might be that $\alpha = 0$. In this case the limiting variance of the residual statistic becomes $\tau^2 + c^T \Gamma c \geq \tau^2$. When $\alpha = 0$, the test statistic $T_n(\theta)$ is asymptotically independent from the estimator $\hat{\theta}_n$ and a test based on estimated residuals always has a larger asymptotic variance than the same test applied to the actual innovations (unless $c = 0$ in which case both variances are equal). The asymptotic independence of the statistic $T_n(\theta)$ and the estimator $\hat{\theta}_n$ implies that the residual-based statistic $T_n(\hat{\theta}_n)$ essentially behaves as a mixture over various values of θ. Such a mixture distribution clearly has a larger variance than the distribution of $T_n(\theta)$ with θ fixed.

3.3 Power of residual-based tests

A question that arises naturally at this point is the effect on the power of a test when applied to residuals rather than innovations. First of all, note that the limiting distribution of the residual-based test statistic in (3.1) clearly does not depend on the local parameter sequence θ_n. This implies that the statistic's distribution is insensitive to local changes in the underlying parameter θ. The test statistic $T_n(\hat{\theta}_n)$, consequently, is valid for the model at hand. However, with the application of specification testing in mind, one may be interested in the local power with respect to other parameters, e.g., of the innovation distribution (like skewness) or omitted variables.

Consider the case where there is an additional parameter ψ in the model and we are interested in the (local) power of the residual-based statistic

$T_n(\hat{\theta}_n)$ with respect to this parameter. The model now consists of a set of probability measures $\{\mathbb{P}^{(n)}_{\theta,\psi}:\ \theta \in \Theta,\ \psi \in \Psi\}$. For ease of notation we assume that the original model is obtained by setting $\psi = 0$, i.e., $\mathbb{P}^{(n)}_{\theta,0} = \mathbb{P}^{(n)}_{\theta}$. As before, fix $\theta_0 \in \Theta$ and consider the local parametrization $(\theta_n, \psi_n) = (\theta_0 + \delta/\sqrt{n}, 0 + \eta/\sqrt{n})$. Introduce the log-likelihood

$$\tilde{\Lambda}^{(n)}(\psi_n|0) = \log \frac{\mathrm{d}\mathbb{P}^{(n)}_{\theta_0,\psi_n}}{\mathrm{d}\mathbb{P}^{(n)}_{\theta_0,0}},$$

with respect to the parameter ψ. We are interested in the behavior of our test statistic $T_n(\hat{\theta}_n)$ under $\mathbb{P}^{(n)}_{\theta_0,\psi_n}$. Assume that Condition (ULAN) is satisfied jointly in θ and ψ. Moreover, assume the equivalent of Condition (AN) under $\psi = 0$, i.e., under $\mathbb{P}^{(n)}_{\theta_n,0}$ and as $n \to \infty$,

$$\begin{aligned}
\begin{bmatrix} T_n(\theta_n) \\ \Lambda^{(n)}(\theta_n|\theta_0) \\ \sqrt{n}\left(\hat{\theta}_n - \theta_n\right) \\ \tilde{\Lambda}^{(n)}(\psi_n|0) \end{bmatrix}
&\xrightarrow{\mathcal{L}}
\begin{bmatrix} T \\ \frac{1}{2}\delta^T I_F \delta + \delta^T \Delta \\ Z \\ -\frac{1}{2}\eta^T I_P \eta + \delta^T I_{FP}\eta + \eta^T \tilde{\Delta} \end{bmatrix} \\
&\sim N\left(\begin{bmatrix} 0 \\ \frac{1}{2}\delta^T I_F \delta \\ 0 \\ -\frac{1}{2}\eta^T I_P \eta + \delta^T I_{FP}\eta \end{bmatrix}, \begin{bmatrix} \tau^2 & c^T\delta & \alpha^T & d^T\eta \\ \delta^T c & \delta^T I_F \delta & \delta^T & \delta^T I_{FP}\eta \\ \alpha & \delta & \Gamma & B^T\eta \\ \eta^T d & \eta^T I_{FP}\delta & \eta^T B & \eta^T I_P \eta \end{bmatrix} \right).
\end{aligned} \tag{3.11}$$

Here I_P denotes the Fisher information for the parameter ψ, while I_{FP} denotes the cross Fisher information between θ and ψ.

The matrix B measures the covariance between the log-likelihood ratio with respect to ψ and the estimator for θ_n. Consequently, this matrix measures the bias in $\hat{\theta}_n$ that occurs due to possible local changes in ψ. The special case $B = 0$ refers to the situation where $\hat{\theta}_n$ is insensitive to local changes in ψ. This occurs, e.g., if $\hat{\theta}_n$ is an efficient estimator for θ in a model where ψ is considered a nuisance parameter. The asymptotic mean of $\tilde{\Lambda}^{(n)}(\psi_n|0)$ in (3.11) is a direct consequence of the fact that the limiting distribution is studied under $(\theta, \psi) = (\theta_n, 0)$. Note that local uniformity with respect to ψ is not required.

The derivations leading to Theorem 3.1 remain valid and can be carried out while taking into account the joint behavior of $T_n(\hat{\theta}_n)$ and $\tilde{\Lambda}^{(n)}(\psi_n|0)$. Under $\mathbb{P}^{(n)}_{\theta_n,0}$, one easily verifies for $(T_n(\hat{\theta}_n), \tilde{\Lambda}^{(n)}(\psi_n|0))$ the following limiting distribution:

$$N(\begin{bmatrix} 0 \\ -\frac{1}{2}\eta^T I_P \eta + \delta^T I_{FP} \eta \end{bmatrix}, \begin{bmatrix} \tau^2 + (\alpha - \Gamma c)^T \Gamma^{-1}(\alpha - \Gamma c) - \alpha^T \Gamma^{-1} \alpha & \eta^T (d - Bc) \\ (d - Bc)^T \eta & \eta^T I_P \eta \end{bmatrix}).$$

Applying Le Cam's third lemma once more, we see that the shift in the innovation-based statistic $T_n(\theta)$ due to local changes in ψ is given by $d^T \eta$, while the same local change in ψ induces a shift of size $(d - Bc)^T\eta$ in the residual-based statistic $T_n(\hat{\theta})$. Consequently, while the local power of the innovation-based static $T_n(\theta)$ is determined by d/τ, that of the residual-based statistic $T_n(\hat{\theta}_n)$ is determined by

$$(d - Bc)/\sqrt{\tau^2 + (\alpha - \Gamma c)^T\Gamma^{-1}(\alpha - \Gamma c) - \alpha^T\Gamma^{-1}\alpha}. \tag{3.12}$$

In case $c = 0$, we find that not only the size of the residual-based statistic is unaltered, but also that its power equals that of the innovation-based statistic. In the special case that $B = 0$, we thus find that the power against local changes in ψ in the residual-based statistic decreases, remains unchanged, or increases as the limiting variance of the residual-based statistic increases, remains unchanged, or decreases, respectively. It may thus very well be the case that residual-based statistics have more power against certain local alternatives than the same statistic applied to actual innovations.

Alternatively, the results in this section can be interpreted in terms of specification testing with locally misspecified alternatives much in the same spirit as Bera and Yoon (1993). Bera and Yoon (1993) derive a correction to standard LM tests which makes them insensitive to local misspecification. Not surprisingly, this correction exactly contains the covariance term B, which is $J_{\psi\phi}$ in their Formula (3.2).

Power of Kendall's tau for the Binary Choice model: For this application, let ψ denote the coefficient of the possibly omitted variable Z. Using (2.2), we find, at $\psi = 0$, $B = \mathbb{E}\varepsilon_i^G(\theta)^2 XZ$ and $d = \mathbb{E}\varepsilon_i^G(\theta)^2 Z^2$. Since we use the maximum likelihood estimator $\hat{\theta}_n^{(ML)}$, we have $\alpha = \Gamma c$ and $\Gamma = I_F^{-1}$. An expression for α was given in (3.5). Consequently, the local power of the residual-based test is determined by the shift

$$\frac{d - BI_F\alpha}{\sqrt{4/9 - \alpha^T I_F \alpha}}.$$

□

Power of runs test for the Dynamic Volatility model: Recall that in this application, we have $c = 0$. As a result, the (local) power of the residual-based runs test for conditional symmetry, equals that of the innovations-based test. □

4 Main result: Formalization

The problem with studying the asymptotic behavior of $T_n(\hat{\theta}_n)$ is that arbitrary estimators $\hat{\theta}_n$ (even if they are regular) can pick out very special points of the parameter space. Without strong uniformity conditions on the behavior of $T_n(\theta)$ as a function of θ, the residual statistic $T_n(\hat{\theta}_n)$ could behave in an erratic way. We solve this problem by discretizing the estimator $\hat{\theta}_n$. This is a well-known technical trick due to Le Cam. We introduce this approach now and study the behavior of the statistic based on the discretized estimated parameter.

The discretized estimator $\bar{\hat{\theta}}_n$ is obtained by rounding the original estimator $\hat{\theta}_n$ to the nearest midpoint in a regular grid of cubes. To be precise, consider a grid of cubes in $\mathbb{R}^k$ with sides of length $d/\sqrt{n}$. We call d the discretization constant. Then $\bar{\hat{\theta}}_n$ is the estimator obtained by taking the midpoint of the cube to which $\hat{\theta}_n$ belongs. To formalize this even further, introduce the function $r : \mathbb{R}^k \to \mathbb{Z}^k$ which arithmetically rounds each of the components of the input vector to the nearest integer. Then, we may write $\bar{\hat{\theta}}_n = dr(\sqrt{n}\hat{\theta}_n/d)/\sqrt{n}$. Our interest lies in the asymptotic behavior of $T_n(\bar{\hat{\theta}}_n)$ for small d. We first study the behavior of $\bar{\hat{\theta}}_n$ in the following lemma.

Lemma 4.1. *Let the discretization constant $d > 0$ be given. Define the "discretized truth" $\bar{\theta}_n = dr(\sqrt{n}\theta_0/d)/\sqrt{n}$. Then $\sqrt{n}(\bar{\hat{\theta}}_n - \bar{\theta}_n)$ is degenerated on $\{dj : \ j \in \mathbb{Z}^k\}$. Moreover, for $\delta_n \to \delta$ as $n \to \infty$, we have*

$$\mathbb{P}^{(n)}_{\bar{\theta}_n+\delta_n/\sqrt{n}} \left\{\sqrt{n}(\bar{\hat{\theta}}_n - \bar{\theta}_n) = dj\right\} \to \mathbb{P}\left\{N(\delta - dj, \Gamma) \in (-\frac{d}{2}\iota, \frac{d}{2}\iota]\right\}, \quad (4.1)$$

where $\iota = (1, 1, \dots, 1)^T \in \mathbb{Z}^k$.

The above lemma is basic to our formal main result that now can be stated. Both proofs can be found in the Appendix C.

Theorem 4.1. *With the notation introduced above and under Conditions (ULAN) and (AN), we have for* $\delta_n \to \delta$ *and as* $n \to \infty$*,*

$$\lim_{d \downarrow 0} \lim_{n \to \infty} \mathbb{P}^{(n)}_{\bar{\theta}_n + \delta_n/\sqrt{n}} \left\{ T_n(\bar{\hat{\theta}}_n) \leq t \right\} = \mathbb{P}\left\{ X \leq t \right\}, \tag{4.2}$$

where

$$X \sim N\left(0, \tau^2 + (\alpha - \Gamma c)^T \Gamma^{-1} (\alpha - \Gamma c) - \alpha^T \Gamma^{-1} \alpha\right).$$

Remark 4.1. In typical applications of the discretization trick, the end result is a statistic whose first-order asymptotics do not depend on the discretization constant d used. Clearly, in such cases taking the limit for $d \downarrow 0$ in (4.2) is not needed. Our assumptions precisely avoid smoothness of the statistic of interest, and therefore do not allow us to make this claim in general. When viewing asymptotic results as an approximation to finite sample distributions, this may be less of an issue as the discretization constant is an auxiliary variable that one prefers to be close to zero in the first place[2]. □

[2]We thank an anonymous referee for pointing this out.

Remark 4.2. As for the informal derivations in Section 3, the above proof is strongly based on a conditioning argument with respect to the value of the estimator $\hat{\theta}_n$, or, more precisely, that of the local estimation error $\sqrt{n}(\hat{\theta}_n - \theta_n)$. This leads one to believe that it is meaningfully possible to derive LAN conditions for conditional distributions, where the conditioning event is the value of some estimation error. The authors of the present paper have, however, not seen any results in this direction. $\square$

5 Conclusions

This paper introduces a novel asymptotic analysis of residuals-based statistics in a Gaussian limiting framework: The models under consideration are assumed to be locally asymptotically normal (LAN), the statistics being studied have limiting normal distributions, and the estimators under consideration are $\sqrt{n}$-consistent and asymptotically normal. In our approach we do not explicitly require any smoothness of the statistics of interest with respect to the nuisance parameter, but we do impose a locally uniform convergence condition that is satisfied for residual-based statistics. We apply this method to derive several new results. For example, we present a new omitted variable specification test for limited dependent variable models and provide its asymptotic distribution. Our method is also useful for deriving the asymptotic distribution of two-step estimators and nonparametric tests.

The method proposed in the paper can be extended in several interesting directions. First of all, while the Gaussian context has many applications for residual-based testing in econometric models, our method essentially builds on Le Cam's third lemma which is not restricted to Gaussian situations. In particular, limiting χ^2 distributions can be handled easily using the same techniques. Also, in case the model of interest is not specified in terms of likelihoods but in terms of moments (like in GMM settings), the same ideas can be applied (see Andreou and Werker, 2009, for further details). Local deviations of the moment conditions can be cast in likelihood terms such that our method still applies. Last but not least, our approach may also represent the foundations for an alternative to the derivation of asymptotic distributions of non-Gaussian statistics as, for instance, in Locally Asymptotically Mixed Normal (LAMN) models.

Acknowledgements

Part of this research was completed when the first author held a Marie Curie fellowship at Tilburg University (MCFI-2000-01645) and while both authors were visiting the Statistical and Applied Mathematical Sciences Institute (SAMSI). The first author acknowledges support of the European Research Council under the European Community FP7/2008-2012 ERC grant 209116. The second author also acknowledges support from Mik. Comments by two anonymous referees, Anil Bera, Christel Bouquiaux, Rob Engle, Eric Ghysels, Lajos Horváth, Nour Meddahi, Bertrand Melenberg, Werner Ploberger, Eric Renault, Enrique Sentana, conference participants at the ESEM 2004, MEG 2004, and NBER time series 2004 conferences, and seminar participants at the London School of Economics, Université de Montréal, and Tilburg University are kindly acknowledged.

A Le Cam's third Lemma

Le Cam's third lemma is discussed in several modern books on asymptotic statistics, e.g., Hájek and Šidák (1967), Le Cam and Yang (1990), Bickel et al. (1993), or van der Vaart (1998). We recall it in its best-known form, i.e., for asymptotically normal distributions. Consider two sequences of probability measures $\left(\mathbb{Q}^{(n)}\right)_{n=1}^{\infty}$ and $\left(\mathbb{P}^{(n)}\right)_{n=1}^{\infty}$ defined on the common measurable spaces $\left(X^{(n)}, \mathcal{X}^{(n)}\right)_{n=1}^{\infty}$. Assume that the log-likelihood ratios

$\Lambda_n = \log \mathrm{d}\mathbb{Q}^{(n)}/\mathrm{d}\mathbb{P}^{(n)}$ satisfy, jointly with some statistic T_n, under $\mathbb{P}^{(n)}$,

$$\begin{pmatrix} T_n \\ \Lambda_n \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} T \\ \Lambda \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ -\frac{\sigma^2}{2} \end{bmatrix}, \begin{bmatrix} \tau^2 & c \\ c & \sigma^2 \end{bmatrix} \right), \tag{A.1}$$

as $n \to \infty$. Le Cam's third lemma then gives the limiting behavior of the statistic T_n under $\mathbb{Q}^{(n)}$. More precisely it states, under $\mathbb{Q}^{(n)}$,

$$T_n \xrightarrow{\mathcal{L}} N(c, \tau^2),$$

as $n \to \infty$. The intuition for this result is based on the fact that a statistic T which is jointly normally distributed with some log-likelihood ratio Λ as in (A.1), has $N(c, \tau^2)$ distribution under the alternative measure. This non-asymptotic version follows trivially from writing down the appropriate densities and likelihood ratios. Le Cam's third lemma takes this result to the limit.

B Sufficient conditions for Condition (AN)

Condition (AN) is required to hold locally uniform, that is under local alternatives $\theta_n = \theta_0 + O(1/\sqrt{n})$. In this appendix we show that, for residual-based statistics, convergence under fixed $\theta_0 \in \Theta$ generally implies this local uniform convergence. We discuss the three components T_n, $\Lambda^{(n)}$, and $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$

in Condition (AN) separately, but, using the Cramér-Wold device, the arguments can easily be combined to prove locally uniform joint convergence.

First, consider the test statistic T_n. Recall that, in our framework, T_n refers to a residual-based statistic used for specification testing. In such case, we can generally write

$$T_n(\theta) = T\left(\varepsilon_1(\theta), \dots \varepsilon_n(\theta)\right), \tag{B.1}$$

where the innovations $\varepsilon_t(\theta)$ are i.i.d., under $\mathbb{P}_\theta^{(n)}$, and T is some given function. It is not excluded that T_n depends on some exogenous variables as well. Assuming appropriate centering and scaling, such statistics often satisfy an asymptotic representation, for given $\theta = \theta_0$,

$$T_n(\theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau\left(\varepsilon_t(\theta_0), \dots, \varepsilon_{t-l}(\theta_0)\right) + o_p(1), \tag{B.2}$$

under $\mathbb{P}_{\theta_0}^{(n)}$, for some function $\tau : \mathbb{R}^l \to \mathbb{R}$ such that a ($l + 1$-dependent) central limit theorem can be applied to (B.2). We now have the following simple but useful result, which is invoked in all applications mentioned in this paper.

Proposition B.1. *Suppose that the statistic $T_n(\theta)$ can be written as in (B.1) and satisfies (B.2). Then, for local alternatives θ_n to θ_0, we have, under $\mathbb{P}_{\theta_n}^{(n)}$,*

$$T_n(\theta_n) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau\left(\varepsilon_t(\theta_n), \dots, \varepsilon_{t-l}(\theta_n)\right) + o_p(1). \tag{B.3}$$

Moreover, the limiting distribution of $T_n(\theta_n)$ *under* $\mathbb{P}^{(n)}_{\theta_n}$ *does not depend on the local alternatives* θ_n *and, thus, equals the limiting distribution of* $T_n(\theta_0)$ *under* $\mathbb{P}^{(n)}_{\theta_0}$*. Thus, Condition (AN) holds with respect to* $T_n(\theta_n)$*.*

Proof. The result is immediate upon noting that

$$\begin{aligned}\mathcal{L}_{\theta_n}&\left(T_n(\theta_n) - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tau\left(\varepsilon_t(\theta_n),\ldots,\varepsilon_{t-l}(\theta_n)\right)\right)\\ = \ \ \mathcal{L}_{\theta_0}&\left(T_n(\theta_0) - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tau\left(\varepsilon_t(\theta_0),\ldots,\varepsilon_{t-l}(\theta_0)\right)\right),\end{aligned}$$

where $\mathcal{L}_\theta$ denotes the distribution under $\mathbb{P}^{(n)}_{\theta}$. Observe that contiguity of θ_n to θ_0 is actually not even needed for this result. □

Remark B.1. The zero mean condition on the limiting distribution of T in Condition (AN) holds for residual-based statistics discussed in Proposition B.1, but is indeed specific to this area of applications. The condition can easily be relaxed. Suppose that the mean of the limiting distribution of $T_n(\theta_n)$ would be $a^T\delta$. Consider, for given θ_0, the auxiliary statistic $\tilde{T}_n(\theta) = T_n(\theta) - a^T\sqrt{n}(\theta - \theta_0)$. The statistic $\tilde{T}_n$, by construction, satisfies Condition (AN) and our main Theorem 3.1 below can be invoked. This idea could also be applied in the analysis of generated or estimated regressors. □

Concerning the locally uniform convergence of the likelihood ratio in Condition (AN), we observe that this is immediately given the required local uniformity in Condition (ULAN). Merely imposing a LAN condition would, together with the local uniformity required in Condition (AN), essentially imply ULAN. In order to be precise about the scope of our results, we impose condition ULAN from the start.

Concerning the estimator $\hat{\theta}_n$, Condition (AN) imposes that the limiting distribution, in particular its mean, of $\sqrt{n}\left(\hat{\theta}_n - \theta_n\right)$, under $\mathbb{P}^{(n)}_{\theta_n}$, does not depend on the local alternatives θ_n and, hence, equals the limit of $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$, under $\mathbb{P}^{(n)}_{\theta_0}$. This is to say that the estimator is regular in the sense of Bickel et al. (1993), page 18, or van der Vaart (1998), page 115. Regularity also implies that the asymptotic covariance between the estimator and the central sequence is the $k \times k$ identity matrix I_k. In particular, the convolution theorem stating that asymptotic variances of estimators are always larger than the inverse of the Fisher information only applies to regular estimators; regularity is used to rule out superefficient estimators. Estimators that satisfy an asymptotically linear representation can often be transformed into regular estimators. This is formalized in the proposition below, whose proof follows easily along the lines of, e.g., van der Vaart (1998), Section 5.7.

Proposition B.2. *Maintaining the Condition (ULAN), consider an estima-*

tor $\hat{\theta}_n$ that satisfies, under $\mathbb{P}^{(n)}_{\theta_0}$, the asymptotic representation

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_t(\theta_0) + o_p(1), \tag{B.4}$$

with ψ_t some influence function that satisfies, still under $\mathbb{P}^{(n)}_{\theta_0}$,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_t(\theta_0) \xrightarrow{\mathcal{L}} N(0,\Gamma).$$

Moreover, assume, for local alternatives θ_n to θ_0,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_t(\theta_n) - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_t(\theta_0) = -\sqrt{n}\,(\theta_n - \theta_0) + o_p(1), \tag{B.5}$$

under $\mathbb{P}^{(n)}_{\theta_0}$. Then, we can construct an estimator $\tilde{\theta}_n$ such that, under $\mathbb{P}^{(n)}_{\theta_n}$,

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_n\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_t(\theta_n) + o_p(1), \tag{B.6}$$

and, under $\mathbb{P}^{(n)}_{\theta_0}$, $\sqrt{n}\left(\tilde{\theta}_n - \hat{\theta}_n\right) = o_p(1)$.

Proof. The construction follows the ideas in, e.g., van der Vaart (1998), Section 5.7. First of all, discretize the estimator $\hat{\theta}_n$ as described in Section 4. Denote this estimator $\bar{\hat{\theta}}_n$. Define the estimator

$$\tilde{\theta}_n := \bar{\hat{\theta}}_n + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_t(\bar{\hat{\theta}}_n). \tag{B.7}$$

Observe, for local alternatives $\overline{\theta}_n$ and by applying (B.5) twice,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_t(\overline{\theta}_n) - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_t(\theta_n) = -\sqrt{n}\left(\overline{\theta}_n - \theta_n\right) + o_p(1). \tag{B.8}$$

Combining (B.7) with (B.8) gives

$$\begin{aligned}\sqrt{n}\left(\tilde{\theta}_n - \theta_n\right) &= \sqrt{n}\left(\bar{\hat{\theta}}_n - \theta_n\right) + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_t(\bar{\hat{\theta}}_n) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_t(\theta_n) + o_p(1),\end{aligned}$$

where the fact that $\bar{\hat{\theta}}_n$ may be considered deterministic follows from its discreteness. See van der Vaart (1998), Section 5.7, or Section 4 in the present paper for details. The statement about $\sqrt{n}\left(\tilde{\theta}_n - \hat{\theta}_n\right)$ follows from (B.4) and (B.6) for $\theta_n = \theta_0$. $\square$

If we forget about the discretization, the estimator introduced in (B.7) actually equals the original estimator $\hat{\theta}_n$ in case

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_t(\hat{\theta}_n) = 0.$$

This is obviously the case for any estimator that exactly solves some appropriate score equations; a Z-estimator. In that respect, modulo the technical discretization issue, any Z-estimator satisfies Condition (AN) as long as (B.5) holds. Condition (B.5) usually follows from the observation that the influence function of an estimator generally satisfies $\mathbb{E}_\theta \frac{\partial}{\partial\theta^T}\psi_t(\theta) = -I_k$, with I_k the identity matrix of dimension $k = dim(\theta)$. Applying a Taylor expansion to (B.5) then motivates the condition. In the particular case of ML estimation, Condition (B.5) follows immediately from Condition (ULAN). Although papers in the econometrics or statistics literature do not always explicitly state or check regularity of proposed estimators, the additional step of proving (B.5) in our case is usually not very complicated and generally does not impose additional regularity conditions.

C Proofs of main results

We first recall the so-called Le Cam's first lemma. For a more detailed discussion we refer to van der Vaart (1998). Le Cam's third lemma essentially states the the operations of taking the limit (as $n \to \infty$) and changing the underlying probability measure from the null to a (local) alternative can be interchanged. More precisely, consider the situation of Condition (AN) with respect to the statistic of interest $T_n(\theta)$ and the log-likelihood ratio $\Lambda^{(n)}(\theta_n|\theta)$. Le Cam's third lemma asserts that the limiting distribution of $T_n(\theta)$ under the local alternatives $\theta_n = \theta + \delta/\sqrt{n}$ is the same as the distribution of T under the change-of-measure induced by the log-likelihood ratio $\frac{1}{2}\delta^T I_F \delta + \delta^T \Delta$ in the limit distribution in Condition (AN). In the ubiquitous case of a Gaussian limit distribution, the resulting limit is well-known to be Gaussian with the same variance τ^2, but with a shift in the mean of size $-c^T\delta$.

Proof of Lemma 4.1: The fact that $\sqrt{n}(\bar{\hat{\theta}}_n - \bar{\theta}_n)$ is degenerated on $\{dj : j \in \mathbb{Z}^k\}$ follows easily from $\sqrt{n}(\bar{\hat{\theta}}_n - \bar{\theta}_n) = dr(\sqrt{n}\hat{\theta}_n/d) - dr(\sqrt{n}\theta_0/d)$. To deduce its limiting distribution, observe the following equalities of events, for fixed $j \in \mathbb{Z}^k$,

$$
\begin{aligned}
\left\{\sqrt{n}(\bar{\hat{\theta}}_n - \bar{\theta}_n) = dj\right\} &= \left\{dr(\sqrt{n}\hat{\theta}_n/d) - dr(\sqrt{n}\theta_0/d) = dj\right\} \\
&= \left\{r(\sqrt{n}\hat{\theta}_n/d) - r(\sqrt{n}\theta_0/d) = j\right\} \\
&= \left\{j - \frac{1}{2}\iota < \frac{\sqrt{n}(\hat{\theta}_n - \bar{\theta}_n)}{d} \le j + \frac{1}{2}\iota\right\} \\
&= \left\{-\frac{d}{2}\iota < \sqrt{n}(\hat{\theta}_n - \bar{\theta}_n) - dj \le \frac{d}{2}\iota\right\}.
\end{aligned}
$$

From the Conditions (ULAN) and (AN), we find, under $\mathbb{P}^{(n)}_{\bar{\theta}_n + dj/\sqrt{n}}$, as $\delta_n \to \delta$, and as $n \to \infty$,

$$
\begin{bmatrix} \Lambda(\bar{\theta}_n + \delta_n/\sqrt{n} | \bar{\theta}_n + dj/\sqrt{n}) \\ \sqrt{n}\left(\hat{\theta}_n - (\bar{\theta}_n + dj/\sqrt{n})\right) \end{bmatrix} \xrightarrow{\mathcal{L}} \begin{bmatrix} -\frac{1}{2}(\delta - dj)^T I_F(\delta - dj) + (\delta - dj)^T\Delta \\ Z \end{bmatrix},
$$

with $[Z, \Delta^T]^T$ as in (3.10). From Le Cam's third lemma, this implies, under $\mathbb{P}^{(n)}_{\bar{\theta}_n + \delta_n/\sqrt{n}}$ and as $n \to \infty$, $\sqrt{n}\left(\hat{\theta}_n - (\bar{\theta}_n + dj/\sqrt{n})\right) = \sqrt{n}\left(\hat{\theta}_n - \bar{\theta}_n\right) - dj \xrightarrow{\mathcal{L}} N(\delta - dj, \Gamma)$. Together with the above result on the event $\left\{\sqrt{n}(\bar{\hat{\theta}}_n - \bar{\theta}_n) = dj\right\}$, the lemma now follows. $\square$

Proof of Theorem 4.1: From the proof of Lemma 4.1, we know

$$\left\{\sqrt{n}(\bar{\hat{\theta}}_n - \bar{\theta}_n) = dj\right\} = \left\{-\frac{d}{2}\iota < \sqrt{n}(\hat{\theta}_n - \bar{\theta}_n) - dj \leq \frac{d}{2}\iota\right\}.$$

Moreover, applying Le Cam's third lemma as in the proof of Lemma 4.1, we find under $\mathbb{P}^{(n)}_{\bar{\theta}_n+\delta_n/\sqrt{n}}$ and as $n \to \infty$,

$$\begin{bmatrix} T_n(\bar{\theta}_n + dj/\sqrt{n}) \\ \sqrt{n}(\hat{\theta}_n - \bar{\theta}_n) - dj \end{bmatrix} \xrightarrow{\mathcal{L}} N\left(\begin{bmatrix} (\delta - dj)^T c \\ \delta - dj \end{bmatrix}, \begin{bmatrix} \tau^2 & \alpha^T \\ \alpha & \Gamma \end{bmatrix}\right).$$

Taking these two results together, we get, for all $j \in \mathbb{Z}^k$ and with the distribution (3.10),

$$\begin{aligned}
&\mathbb{P}^{(n)}_{\bar{\theta}_n+\delta_n/\sqrt{n}}\left\{T_n(\bar{\theta}_n + dj/\sqrt{n}) \leq t \text{ and } \sqrt{n}(\bar{\hat{\theta}}_n - \bar{\theta}_n) = dj\right\} \\
&\quad \to \quad \mathbb{P}\left\{T + (\delta - dj)^T c \leq t \text{ and } -\frac{d}{2}\iota < Z + (\delta - dj) \leq \frac{d}{2}\iota\right\}.
\end{aligned}$$

The number of values that $\sqrt{n}(\bar{\hat{\theta}}_n - \bar{\theta}_n)$ takes in a bounded set, is finite. Consequently, we may write for each $M > 0$,

$$\begin{aligned}
&\mathbb{P}^{(n)}_{\bar{\theta}_n+\delta_n/\sqrt{n}}\left\{T_n(\bar{\hat{\theta}}_n) \leq t \text{ and } \left|\sqrt{n}(\bar{\hat{\theta}}_n - \bar{\theta}_n)\right| \leq M\right\} \\
&\quad = \sum_{j\in\mathbb{Z}^k,\ d|j|\leq M} \mathbb{P}^{(n)}_{\bar{\theta}_n+\delta_n/\sqrt{n}}\left\{T_n(\bar{\theta}_n + dj/\sqrt{n}) \leq t \text{ and } \sqrt{n}(\bar{\hat{\theta}}_n - \bar{\theta}_n) = dj\right\} \\
&\quad \to \sum_{j\in\mathbb{Z}^k,\ d|j|\leq M} \mathbb{P}\left\{T + (\delta - dj)^T c \leq t \text{ and } -\frac{d}{2}\iota < Z + (\delta - dj) \leq \frac{d}{2}\iota\right\},
\end{aligned}$$

as $n \to \infty$. Since $\limsup_{n\to\infty} \mathbb{P}^{(n)}_{\bar{\theta}_n+\delta_n/\sqrt{n}} \left\{ \left| \sqrt{n}(\bar{\hat{\theta}}_n - \bar{\theta}_n) \right| > M \right\} \to 0$ as $M \to \infty$, we obtain

$$\begin{aligned}
&\mathbb{P}^{(n)}_{\bar{\theta}_n+\delta_n/\sqrt{n}} \left\{ T_n(\bar{\hat{\theta}}_n) \le t \right\} \\
&\quad \to \sum_{j\in\mathbb{Z}^k} \mathbb{P} \left\{ T \le t - (\delta - dj)^T c \text{ and } -\frac{d}{2}\iota < Z + (\delta - dj) \le \frac{d}{2}\iota \right\},
\end{aligned}$$

as $n \to \infty$. Let φ_{TZ} denote the probability density function of $[T, Z^T]^T$ and φ_Z that of Z. Observe that, conditionally on $Z = z$, $T \sim N(\alpha^T\Gamma^{-1}z, \tau^2 - \alpha^T\Gamma^{-1}\alpha)$. Consequently,

$$\begin{aligned}
&\sum_{j\in\mathbb{Z}^k} \mathbb{P} \left\{ T \le t - (\delta - dj)^T c \text{ and } -\frac{d}{2}\iota < Z + (\delta - dj) \le \frac{d}{2}\iota \right\} \\
&= \sum_{j\in\mathbb{Z}^k} \int_{x=-\infty}^{t-(\delta-dj)^T c} \int_{z=-(\delta-dj)-\frac{d}{2}\iota}^{-(\delta-dj)+\frac{d}{2}\iota} \varphi_{TZ}(x,z)\mathrm{d}x\mathrm{d}z \\
&= \sum_{j\in\mathbb{Z}^k} \int_{z=-(\delta-dj)-\frac{d}{2}\iota}^{-(\delta-dj)+\frac{d}{2}\iota} \Phi\left(\frac{t - (\delta - dj)^T c - \alpha^T\Gamma^{-1}z}{\sqrt{\tau^2 - \alpha^T\Gamma^{-1}\alpha}} \right) \varphi_Z(z)\mathrm{d}z \\
&= \sum_{j\in\mathbb{Z}^k} \int_{z=-(\delta-dj)-\frac{d}{2}\iota}^{-(\delta-dj)+\frac{d}{2}\iota} \Phi\left(\frac{t - (\alpha - \Gamma c)^T\Gamma^{-1}z}{\sqrt{\tau^2 - \alpha^T\Gamma^{-1}\alpha}} \right) \varphi_Z(z)\mathrm{d}z + O(d) \\
&= \int_{z\in\mathrm{R}^k} \Phi\left(\frac{t - (\alpha - \Gamma c)^T\Gamma^{-1}z}{\sqrt{\tau^2 - \alpha^T\Gamma^{-1}\alpha}} \right) \varphi_Z(z)\mathrm{d}z + O(d) \\
&= \int_{z\in\mathrm{R}^k} \mathbb{P}\left\{ X \le t | Z = z \right\} \varphi_Z(z)\mathrm{d}z + O(d) \\
&\to \mathbb{P}\left\{ X \le t \right\},
\end{aligned}$$

as $d \downarrow 0$, with

$$\begin{bmatrix} X \\ Z \end{bmatrix} \sim N\left(0, \begin{bmatrix} \tau^2 + (\alpha - \Gamma c)^T \Gamma^{-1}(\alpha - \Gamma c) - \alpha^T \Gamma^{-1} \alpha & \alpha - \Gamma c \\ (\alpha - \Gamma c)^T & \Gamma \end{bmatrix}\right).$$

$\square$

References

Abadie A. and G.W. Imbens (2008). On the failure of the bootstrap for matching estimators. *Econometrica* **76**, **6**, 1537-1558.

Abadie A. and G.W. Imbens (2009). Matching on the estimated propensity score. *NBER Working Paper*.

Abadir K. and W. Distaso (2007). Testing joint hypotheses when one of the alternatives is one-sided. *Journal of Econometrics* **140**, 695-718.

Andreou, E., and B.J.M. Werker (2009). An Alternative Asymptotic Analysis of Residual-Based Statistics. *Tilburg University Working Paper*.

Andrews, D.W.K. (1994). Empirical process methods in econometrics. *Handbook of Econometrics* **37**.

Bai, J., and S. Ng (2001). A consistent test for symmetry in times series models. *Journal of Econometrics* **103**, 225–158.

Bera, A.K., and G. Premaratne (2005). A test for symmetry with leptokurtic financial data. *Journal of Financial Econometrics* **3**, 169–187.

Bera, A.K., and G. Premaratne (2009). Adjusting the tests for skewness and kurtosis for distributional misspecifications. Working paper.

Bera, A.K., and M.J. Yoon (1993). Specification testing with misspecified local alternatives. *Econometric Theory* **9**, 649–658.

Berkes, I., and L. Horváth (2004). The efficiency of the estimator of the parameters in GARCH processes. *The Annals of Statistics* **32**, 633–655.

Bickel, P.J., C.A.J. Klaassen, Y. Ritov, and J.A. Wellner (1993). *Efficient and Adaptive Statistical Inference for Semiparametric Models.* John Hopkins University Press, Baltimore.

COTTER, K.D., (1986). Similarity of information and behavior with a pointwise convergence topology. *Journal of Mathematical Economics* **15**, 25–38.

DROST, F.C. and C.A.J. KLAASSEN (1997). Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* **81**, 193–221.

HALLIN, M. and M.L. PURI (1991). Time series analysis via rank-order theory: signed-rank tests for ARMA models. *Journal of Multivariate Analysis* **39**, 175–237.

JEGANATHAN, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* **11**, 818–887.

LE CAM, L. and G.L. YANG (1990). *Asymptotics in Statistics; Some Basic Concepts*. Springer, Berlin.

LINTON, O. (1993). Adaptive estimation in ARCH models. *Econometric Theory* **9**, 539–569.

LUMSDAINE, R. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* **64**, 575–596.

Murphy, K., and R. Topel (1985). Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics* **3**, 370–379.

Murphy, K., and R. Topel (2002). Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics* **20**, 88–97.

Nelson, D. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6**, 318–334.

Newey, W., and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, Ch. 36, Vol. IV, R. Engle and D. McFadden (eds.).

Pagan, A. (1986). Two Stage and Related Estimators and Their Applications. *The Review of Economic Studies* **53**, 517–538.

Pagan, A., and F. Vella (1989). Diagnostic tests for models based on individual data: A Survey. *Journal of Applied Econometrics* **4**, S29–S59.

Pierce, D.A. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *The Annals of Statistics* **10**, 475–478.

Ploberger, W. (2004). A complete class of tests when the likelihood is locally asymptotically quadratic. Journal of Econometrics, 118, 67-94

Pollard, D. (1989). Asymptotics via empirical processes, *Statistical Science* **4**, 341–366.

Randles, R.H. (1982). On the asymptotic normality of statistics with estimated parameters. *The Annals of Statistics* **10**, 462–474.

Strasser, H. (1985). *Mathematical Theory of Statistics.* Walter de Gruyter, New York.

Sun, Y., and T. Tsengos (2006). Semiparametric efficient adaptive estimation of asymmetric GARCH models. *Journal of Econometrics* **133**, 373–386.

van der Vaart, A. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.