A Conditional Goodness-of-Fit Test for Time Series

Atsushi Inoue *

First Version: June 1997

This Version: July 1998

Abstract

This paper proposes a unified approach for consistent testing of linear restrictions on the conditional distribution function of a time series. A wide variety of interesting hypotheses in economics and finance correspond to such restrictions, including hypotheses involving conditional goodness-of-fit, conditional homogeneity, conditional mixtures, conditional quantiles, conditional symmetry, distributional Granger non-causality, and interval forecasts. The finite-sample properties are investigated in a set of Monte Carlo experiments. The proposed tests are conservative but perform well in samples of the size relevant for empirical finance.

Key Words: conditional distribution functions, consistent testing, time series.

*My thanks go to F.X. Diebold for his advice and support, to V. Corradi, J. Hahn, L. Kilian, R.S. Mariano, N. Meddahi, P.E. Todd, and seminar participants at Cornell, Hitotsubashi, North Carolina State, Penn, Rochester, and Yale for helpful comments. and to R.C. Douglas for providing me excellent computer facilities. Of course, any remaining errors are my own responsibility. From August on, address correspondence to: Atsushi Inoue, Department of Agricultural and Resource Economics, Box 8109, North Carolina State University, Raleigh, NC 27695-8109. E-mail: inoue@ssc.upenn.edu.

1. Introduction

This paper proposes a unified approach for consistent testing of linear restrictions on the conditional distribution function (d.f.) of a time series. A wide variety of interesting and important hypotheses in economics and finance correspond to such restrictions, including hypotheses involving conditional goodness-of-fit, conditional homogeneity, conditional quantiles, conditional symmetry, distributional Granger non-causality, and interval forecasts. The tests, for example, are naturally-suited to helping answer questions such as "Are the distributions of assets, consumption, and income implied by a particular dynamic macroeconomic model close to the actual distributions in the data?," or "Do differences in microstructure across different asset markets produce differences in return distributions, and if so, are the differences consistent with a particular microstructure theory?"

The econometric literature on conditional d.f.s is growing rapidly. For example, specification tests for the parametric conditional d.f. for iid observations have been considered by Heckman (1984), Andrews (1988b, 1997), Stinchcombe and White (1995), and Zheng (1994, 1996). Moreover, those for dependent observations have been considered recently by Diebold, Gunther, and Tay (1997) and Bai (1997). This paper differs from the literature in that it considers consistent tests for a much wider range of restrictions on the conditional d.f.s in a unified framework. Moreover, this paper allows for the possibility of an infinite history of information, which is important because many economic and financial time series may not be Markovian. The proposed test is consistent against all alternatives to the null hypothesis.

The remainder of the paper is organized as follows. Section 2 states the canonical form of the null hypothesis and lists a variety of hypotheses in economics and finance which imply linear restrictions on the conditional d.f.s. Section 3 develops an asymptotic theory. Section 4 assesses the finite-sample performance of the proposed test for conditional goodness-of-fit restrictions in a set of Monte Carlo experiments. We find that the proposed test is conservative but powerful enough for the sample size relevant for empirical finance. An appendix provides proofs of the results given in the paper.

2. Hypothesis of Interest

This section states the canonical form of linear restrictions on the conditional d.f.s and shows that a wide variety of interesting hypotheses in econometrics imply linear restrictions on the conditional d.f.s.

Let $\{z_t\}_{t=-\infty}^{\infty}$ be a strictly stationary sequence of $\Re^d$-valued random variables. Let $u_t(\theta) = g(z_t, \theta)$ where $g : \Re^d \times \Theta \to \Re^r$, $\Theta \subset \Re^p$, and let $F_j(\cdot|\mathcal{Z}_{t-m}^{t-1}, \theta)$ denote the d.f. of $u_{tj}(\theta)$ conditional on $z_{t-1}, z_{t-2}, ..., z_{t-m}$. Let $c = (c_1, c_2, ..., c_r) \in \Re^r$, $\upsilon = (\upsilon_1, \upsilon_2, ..., \upsilon_r) \in \Upsilon \subset \Re^r$, and $G(\upsilon|\mathcal{Z}_{-\infty}^{t-1}, \theta)$ be measurable with respect to $\mathcal{Z}_{-\infty}^{t-1}$. As shown later, c, Υ, and $G(\upsilon|\mathcal{Z}_{-\infty}^{t-1}, \theta_0)$ are specified as part of the null hypothesis. The canonical form of the null hypothesis is that there exists some $\theta_0 \in \Theta$ such that, for all $\upsilon \in \Upsilon$,

$$P\Big(\sum_{j=1}^{r} c_j F_j(\upsilon_j|\mathcal{Z}_{-\infty}^{-1}, \theta_0) = G(\upsilon|\mathcal{Z}_{-\infty}^{-1}, \theta_0)\Big) = 1. \tag{2.1}$$

In other words, the linear combination of conditional d.f.s is equal to the function G under the null hypothesis.[1] The alternative hypothesis is the negation of the null hypothesis (2.1).

[1] Note that the null hypothesis does not require the uniqueness of θ_0 but the existence of θ_0. The case of set-valued θ_0 will be mentioned in Section 3.

That is, for every $\theta \in \Theta$, there exists some $v \in \Upsilon$ such that

$$P\Big(\sum_{j=1}^{r} c_j F_j(v_j|\mathcal{Z}_{-\infty}^{-1}, \theta) = G(v|\mathcal{Z}_{-\infty}^{-1}, \theta) \Big) < 1. \tag{2.2}$$

(2.1) can accommodate and unify a wide variety of interesting hypotheses that arise in econometric applications as shown below.

2.1. Conditional Goodness of Fit

A conditional goodness-of-fit restriction is

$$P(y_t \leq v|z_{t-1}, z_{t-2}, ...) = G(v|\mathcal{Z}_{-\infty}^{t-1}, \theta_0) \text{ a.s.} \tag{2.3}$$

for all v, where $G(\cdot|\cdot,\cdot)$ is a parametric conditional d.f. That is, the conditional d.f. of y_t given z_{t-1}, z_{t-2},... is in the parametric family $\{G(v|\mathcal{Z}, \theta) : \theta \in \Theta\}$. This problem has been considered by Heckman (1984), Andrews (1988b, 1997), Stinchcombe and White (1995), and Zheng (1994,1996) in the iid context, and by Diebold, Gunther, and Tay (1997) and Bai (1997) in the time-series context. Letting $r = 1$ and $c_1 = 1$, one can express the conditional goodness-of-fit restriction (2.3) in the form of (2.1).

The conditional goodness-of-fit restriction can be interesting in economics and finance. The parametric d.f. $G(\cdot|\cdot,\cdot)$ can be theoretically derived or simulated from a particular economic model. In macroeconomics, one can test the specification of a dynamic economic model by testing the conditional goodness-of-fit restriction. For example, one may test whether the distributions of assets, consumptions, and income implied by a particular dynamic economic model are close to the actual distributions in the data. Diebold *et al.* (1997) show that integral transformed random variables are iid uniform under the correct specification, and they use the result to evaluate the density forecasts of GARCH models.

Bai (1997) improves upon the test of Diebold *et al.* (1997) by explicitly taking account of estimation uncertainty, and he also considers GARCH models as a special case of his general theory.

2.2. Conditional Homogeneity

A conditional homogeneity restriction is

$$F_1(v|\mathcal{Z}_{-\infty}^{t-1}) = F_2(v|\mathcal{Z}_{-\infty}^{t-1}) \tag{2.4}$$

for all v, where F_1 and F_2 are the conditional d.f.s of u_1 and u_2, respectively. Letting $r = 2$, $c_1 = 1$, $c_2 = -1$, $G(\cdot|\cdot, \theta_0) = 0$, and $\Upsilon = \{(v_1, v_2) \in \Re^2 : v_1 = v_2\}$, we can write (2.4) in the form of (2.1).

While tests for unconditional homogeneity have been well studied in the statistical literature, tests for conditional homogeneity have not. Nonetheless, conditional homogeneity can be interesting in various contexts of economics and finance where dynamics and conditioning are often important. In financial economics, one might test the difference in market microstructure by testing the equality of conditional d.f.s. between the exchange rates in New York and in London. In labor economics, one may test the difference in wage distributions between blacks and whites conditional on age, education, and experience.

2.3. Conditional Quantiles

A conditional quantile restriction is the specification of a quantile function and can be written as

$$P(y_t \leq q_p(x_t, \theta_0)|\mathcal{Z}_{-\infty}^{t-1}) = p \text{ a.s.} \tag{2.5}$$

for some function $q_p(\cdot,\cdot)$ and $p \in (0,1)$. The conditional quantile restriction (2.5) is closely related to the conditional calibration restriction (2.8). While the former restriction is used for estimation, the latter is motivated by forecasting. When $q_p(x,\theta_0) = \theta_0' x$, this restriction is the quantile regression of Koenker and Bassett (1978). Bierens and Ginther (1997) consider a consistent specification test of median regression models for iid observations. Zheng (1998a) develops a kernel-based consistent specification test of quantile regression models for iid observations. While Bierens and Ginther (1997) do not take into account estimation uncertainty, Zheng's (1998a) test and ours explicitly do. Let $r = 1$, $c_1 = 1$, $\Upsilon_1 = \{0\}$, $u_t(\theta) = y_t - q_p(x_t,\theta)$, and $G(0|\cdot,\theta_0) \equiv p$. Then (2.5) can be stated in the form of (2.1).

2.4. Conditional Symmetry

A conditional symmetry (around zero) restriction is

$$P(y_t - f(x_t,\theta_0) \leq -v|\mathcal{Z}_{-\infty}^{t-1}) = P(y_t - f(x_t,\theta_0) \geq v|\mathcal{Z}_{-\infty}^{t-1}) \tag{2.6}$$

for all $v \in \Re$. Fan and Gencay (1995) consider a consistent test for symmetry in linear regression models for iid observations. Using a kernel method, Zheng (1998b) develops a consistent test for symmetry in nonlinear models for iid observations. Let $r = 2$, $c_1 = 1$, $c_2 = -1$, $\Upsilon = \{(v_1,v_2) \in \Re^2 : v_1 = v_2\}$, $u_{t1}(\theta) = y_t - f(x_t,\theta)$, $u_{t2}(\theta) = f(x_t,\theta) - y_t$, and $G(\cdot|\cdot,\theta_0) \equiv 0$. Then (2.6) can be written in the form of (2.1).

Symmetry of the disturbance plays an important role in adaptive estimation (see Bickel, 1982 and Newey, 1988), and quasi-maximum likelihood estimation (see Lumsdaine, 1996, and Newey and Steigerwald, 1997).

2.5. Distributional Granger Non-causality

We say that y_t does not Granger-cause x_t in the distributional sense if

$$P(x_t \leq v | x_{t-1}, x_{t-2}, ..., y_{t-1}, y_{t-2}, ...) = P(x_t \leq v | x_{t-1}, x_{t-2}, ...) \text{ a.s.} \tag{2.7}$$

for all $v \in \Re$. While the conventional Granger non-causality restriction is stated in terms of conditional mean, distributional Granger non-causality is in terms of conditional distribution. An economic model may imply that no additional economic variable can improve its prediction. By testing the distributional Granger non-causality, one can test such hypotheses. Linton and Gozalo (1996) consider a nonparametric test for conditional independence which is closedly related to the concept of distributional Granger non-causality. Let $r = 1$, $c_1 = 1$, $G(v|\cdot, \theta_0) = P(x_t \leq v | x_{t-1}, x_{t-2}, ...)$. Then the Granger non-causality restriction (2.7) can take the form of (2.1). In our framework, the specification of $P(x_t \leq v | x_{t-1}, x_{t-2}, ...)$ must be provided.

2.6. Interval Forecasts

Let $(\underline{y}_\alpha(\cdot, \theta_0), \overline{y}_\alpha(\cdot, \theta_0))$ denote a $100(1-\alpha)\%$ interval forecast of y_t. The interval forecast is correct if and only if

$$P(\underline{y}_\alpha(x_t, \theta_0) \leq y_t \leq \overline{y}_\alpha(x_t, \theta_0) | z_{t-1}, z_{t-2}, ...) = 1 - \alpha, \text{ a.s.} \tag{2.8}$$

Let $r = 2$, $c_1 = 1$, $c_2 = -1$, $\Upsilon_1 = \Upsilon_2 = \{0\}$, $u_{t1}(\theta) = y_t - \overline{y}_\alpha(x_t, \theta)$, $u_{t2}(\theta) = y_t - \underline{y}_\alpha(x_t, \theta)$, $G(0|\cdot, \theta_0) \equiv 1 - \alpha$, and $z_t = (x_{t-1}, y_t)$. Then (2.8) can be expressed in the form of (2.1). While we present the results only for one-step-ahead forecasts, one can always consider multi-step-ahead forecasts at the expense of additional notation for a forecast horizon.

While a point forecast predicts a single future realization of a particular random variable, an interval forecast predicts an area in which the random variable falls with a specified probability. Interval forecasts are becoming important in the literature on forecasting, especially in the context of risk management. Christoffersen (1997) shows that $\{I(\underline{y}_\alpha(x_t,\theta_0) \leq y_t \leq \overline{y}_\alpha(x_t,\theta_0))\}$ is a sequence of iid Bernoulli random variables under the correct specification, and he examines the interval forecasts of GARCH models. While he does not take into account estimation uncertainty due to replacement for the unknown parameter θ_0 by an estimate $\hat{\theta}_T$, we explicitly do. One can also test the specification of a particular econometric model by testing whether it produces correct interval forecasts.

3. Asymptotic Theory

Let $\{\Xi_j\}_{j=1}^{\infty}$ be a sequence of compact subsets of $\Re^d$, $\xi = (\xi_1, \xi_2, ...) \in \Xi_1 \times \Xi_2 \times \cdots = \Xi$, $\gamma = (\xi, \upsilon)$, and $\Gamma = \Xi \times \Upsilon$. Given measurable functions $w : \Re \to \Re$ and $\phi : \Re^d \to \Re^d$, the null hypothesis (2.1) implies

$$E\left(w(\sum_{j=1}^{\infty} \xi_j \phi(z_{-j}))[\sum_{j=1}^{r} c_j F_j(\upsilon_j | \mathcal{Z}_{-\infty}^{-1}, \theta_0) - G(\upsilon | \mathcal{Z}_{-\infty}^{-1}, \theta_0)] \right) = 0 \qquad (3.1)$$

for all $\gamma \in \Gamma$. Because the null (2.1) is equivalent to the unconditional moment restriction (3.1) under certain conditions on $w(\cdot)$ (see Bierens, 1990, and Stinchcombe and White, 1998), we consider consistent testing of (2.1) based on (3.1). Bierens (1990) is the first to consider consistent tests based on the weight function $w(\cdot) = \exp(\cdot)$ in testing for the conditional moment restriction of a nonlinear regression model. The discontinuity of the indicator function renders the existing results inapplicable to our problem, however.

We introduce the following notation.

$$G_v(v|\mathcal{Z}_0^{t-1},\theta) = \partial G(v|\mathcal{Z}_0^{t-1},\theta)/\partial v, \qquad G_\theta(v|\mathcal{Z}_0^{t-1},\theta) = \partial G(v|\mathcal{Z}_0^{t-1},\theta)/\partial \theta,$$

$$G_{vv}(v|\mathcal{Z}_0^{t-1},\theta) = \partial^2 G(v|\mathcal{Z}_0^{t-1},\theta)/\partial v \partial v', \quad G_{v\theta}(v|\mathcal{Z}_0^{t-1},\theta) = \partial^2 G(v|\mathcal{Z}_0^{t-1},\theta)/\partial v \partial \theta',$$

$$g_{i\theta}(z_t,\theta) = \partial g_i(z_t,\theta)/\partial\theta, \qquad g_{i\theta\theta}(z_t,\theta) = \partial^2 g_i(z_t,\theta)/\partial\theta\partial\theta',$$

$$H_v(v|\mathcal{Z}_0^{t-1}) = \partial H(v|\mathcal{Z}_0^{t-1})/\partial v, \qquad H_{vv}(v|\mathcal{Z}_0^{t-1}) = \partial^2 H(v|\mathcal{Z}_0^{t-1})/\partial v \partial v',$$

where $g(z_t,\theta) = u_t(\theta)$ and $H(v|\mathcal{Z}_0^{t-1}) = \sum_{j=1}^r c_j F_j(v_j|\mathcal{Z}_0^{t-1},\theta_1)$ a.s. where $\theta_1 \in \Theta_1 \subset \Re^p$. When θ_0 is set-valued, let θ_0 denote one of such θ_0's from now on. It is possible because estimation is not necessarily based on (2.1). Let $\hat\theta_T$ denote an estimate of θ_0, and $\Theta_0 \subset \operatorname{int}\Theta$ some neighborhood of θ_0.

We introduce the following set of assumptions.

Assumption 1: Under the null hypothesis (2.1),

(a) $\{z_t\}_{t=-\infty}^{\infty}$ is a strictly stationary α-mixing sequence of $\Re^d$-valued random variables with mixing coefficients of size $-\eta/(\eta-2)$ for some $\eta > 2$.

(b) $\{G^2(v|\mathcal{Z}_0^{t-1},\theta_0)\}_{t=1}^\infty$ is L^2-NED on $\{z_t\}$ of size $-1/2$ and L^η-integrable. $\{\|G(v|\mathcal{Z}_{-t}^{-1},\theta_0) - G(v|\mathcal{Z}_{-\infty}^{-1},\theta_0)\|_2\}_{t=1}^\infty$ and $\{\|G_v(v|\mathcal{Z}_{-t}^{-1},\theta_0) - G_v(v|\mathcal{Z}_{-\infty}^{-1},\theta_0)\|_2\}_{t=1}^\infty$ are of size $-1/2$. There exists a sequence of Borel measurable functions $\{\bar G_t(z_{t-1},...,z_0)\}_{t=1}^\infty$ that is L^2-NED of size $-1/2$ on $\{z_t\}$, $L^{2\eta}$-integrable, and satisfies

$$|G_{v_i}(v|\mathcal{Z}_0^{t-1},\theta)| \le \bar G_t(z_{t-1},z_{t-2},...,z_0) \text{ a.s.} \tag{3.2}$$

$$|G_{v_i v_j}(v|\mathcal{Z}_0^{t-1},\theta)| \le \bar G_t(z_{t-1},z_{t-2},...,z_0) \text{ a.s.} \tag{3.3}$$

$$|G_{v_i \theta_k}(v|\mathcal{Z}_0^{t-1},\theta)| \le \bar G_t(z_{t-1},z_{t-2},...,z_0) \text{ a.s.} \tag{3.4}$$

$$|G_{\theta_k}(v|\mathcal{Z}_0^{t-1},\theta)| \le \bar G_t(z_{t-1},z_{t-2},...,z_0) \text{ a.s.} \tag{3.5}$$

for all $\upsilon \in \Upsilon$, $\theta \in \Theta_0$, $i, j = 1, 2, ..., r$, $k = 1, 2, ..., p$, and $t = 1, 2, ...$

(c) $g(z_t, \theta)$ is Borel measurable for all $\theta \in \Theta$. There exists a Borel measurable function $\bar{g}(z_t)$ that is $L^{2\eta}$-integrable and satisfies

$$|g_{i\theta_j}(z_t, \theta)| \leq \bar{g}(z_t) \text{ a.s.} \tag{3.6}$$

$$|g_{i\theta_j\theta_k}(z_t, \theta)| \leq \bar{g}(z_t) \text{ a.s.} \tag{3.7}$$

for $i = 1, 2, ..., r$, $j, k = 1, 2, ..., p$, and $\theta \in \Theta_0$.

(d) $\hat{\theta}_T$ has a linear expansion such that

$$T^{\frac{1}{2}}(\hat{\theta}_T - \theta_0) = T^{-\frac{1}{2}} \sum_{t=1}^{T} \psi_t(z_t, z_{t-1}, ..., z_0, \theta_0) + o_p(1), \tag{3.8}$$

where $\psi_t : \Re^{d(t+1)} \times \Theta \to \Re^p$ is Borel measurable for all $\theta \in \Theta$ and $L^{2\eta}$-integrable, and

$$E[\sum_{t=1}^{T} \psi_t(z_t, z_{t-1}, ..., z_0, \theta_0)|\mathcal{Z}_0^{T-1})] = \sum_{t=1}^{T-1} \psi_t(z_t, z_{t-1}, ..., z_0, \theta_0) \text{ a.s.} \tag{3.9}$$

for all $T = 2, 3, ...$ There exists a sequence of Borel measurable functions $\{\bar{\psi}_t(z_t, ..., z_0)\}_{t=1}^{\infty}$ that is L^2-NED of size $-1/2$ on $\{z_t\}$, $L^{2\eta}$-integrable and satisfies

$$\left| \frac{1}{T} \sum_{t=1}^{T} \{\psi_t(z_t, ..., z_0, \theta') - \psi_t(z_t, ..., z_0, \theta)\} \right|^2 \leq \frac{1}{T} \sum_{t=1}^{T} \bar{\psi}_t(z_t, ..., z_0)|\theta' - \theta| + o_p(1) \tag{3.10}$$

for all $i = 1, 2, ..., p$, $\theta, \theta' \in \Theta_0$, and $t = 1, 2, ...$

(e) $\Upsilon \subset \Re^r$, $\Xi_j \subset \Re^d$ is compact, $\xi_{\max j} = \max_{\xi_j \in \Xi_j} \xi_j = O(j^{-1-\zeta})$, and $\xi_{\min j} = \min_{\xi_j \in \Xi_j} \xi_j = O(j^{-1-\zeta})$ for some $\zeta > 1/2$.

(f) $w : \Re \to \Re$ is non-polynomial analytic, and $\phi : \Re^d \to \Re^d$ is bounded, Borel measurable, and one-to-one.

Under the alternative hypothesis, θ_0 and Θ_0 in the above assumptions are replaced by θ_1 and Θ_1, respectively, and (3.8) and (3.9) are replaced by

(d')

$$T^{\frac{1}{2}}(\hat{\theta}_T - \theta_1) = O_p(1). \tag{3.11}$$

In addition, we assume that

(g) $\{H^2(v|\mathcal{Z}_0^{t-1})\}_{t=1}^{\infty}$ is L^2-NED on $\{z_t\}$ of size $-1/2$ and L^{η}-integrable. $\{\|H(v|\mathcal{Z}_{-t}^{-1}) - H(v|\mathcal{Z}_{-\infty}^{-1})\|_2\}_{t=1}^{\infty}$ is of size $-1/2$. There exists a sequence of Borel measurable functions $\{\bar{H}_t(z_{t-1}, z_{t-2}, ..., z_0)\}_{t=1}^{\infty}$ that is L^2-NED of size $-1/2$ on $\{z_t\}$, $L^{2\eta}$-integrable, and satisfies

$$|H_{v_i}(v|\mathcal{Z}_0^{t-1})| \leq \bar{H}_t(z_{t-1}, z_{t-2}, ..., z_0) \text{ a.s.} \tag{3.12}$$

$$|H_{v_i v_j}(v|\mathcal{Z}_0^{t-1})| \leq \bar{H}_t(z_{t-1}, z_{t-2}, ..., z_0) \text{ a.s.} \tag{3.13}$$

for all $v \in \Upsilon$, $i, j = 1, 2, ..., r$, $t = 1, 2, ...$

(h) There exists a $\Re^d$-valued strictly stationary sequence $\{v_t\}_{t=1}^{\infty}$ such that $G(v|\mathcal{Z}_0^{t-1}, \theta_1) = \sum_{j=1}^{r} c_j P(v_{tj} \leq v_j|\mathcal{Z}_0^{t-1})$ a.s.

Before we present our theorem, we will remark on Assumption 1. The conditions on dependence and moments in Assumptions 1(a)(b)(d)(g) have been commonly used in the literature on nonlinear econometric models; see Gallant and White (1988). (3.2), (3.3), (3.4), (3.5), (3.6), (3.7), (3.12), and (3.13) impose a version of the global Lipschitz condition. Without loss of generality, the same bound functions are used to simplify notation. It is not restrictive that $u_t(\theta) = g(z_t, \theta)$ depends only on z_t. By expanding the dimension of z_t,

$u_t(\theta)$ can depend on more complex dynamics. Assumption 1(d) requires that the estimator $\hat{\theta}_T$ satisfy the CLT for martingales under the null hypothesis. For example, the ordinary least squares (OLS) estimator with martingale-difference disturbances and the maximum likelihood estimator (MLE) based on $\log L(\theta) = \sum_{t=1}^{T} \log f(z_t|z_{t-1}, z_{t-2}, ..., z_0, \theta)$ satisfy (3.8) and (3.9) as long as some conditions on the dependence and moments are satisfied. GMM estimators based on conditional moment restrictions, such as Euler equations, satisfy (3.8) and (3.9). Andrews (1997) also requires the estimator to have a linear expansion. Assumption 1(d') requires that the estimator converges in probability under the alternative hypothesis.

Assumption 1(e) gives conditions on the space of nuisance parameters which is under the control of econometricians. The condition on $\{\Xi_j\}$ in Assumption 1(f) guarantees that the summand in the function w is well-defined (c.f., De Jong (1996, equation 12) who uses $\Xi_j = [a_j, b_j]$ where $a_j, b_j = O(j^{-2})$). The first part of Assumption 1(f) is from Theorem 2.3 in Stinchcombe and White (1995) and guarantees the test consistency against all alternative hypotheses. θ_1 is the limit of $\hat{\theta}_T$ under the alternative hypothesis. For example, if $\hat{\theta}_T$ is a MLE, θ_1 is the limit of the quasi-MLE under the alternative hypothesis.

Let

$$q_{1t}(\gamma) = w(\sum_{j=0}^{t} \xi_j \phi(z_{t-j}))[\sum_{j=1}^{r} c_j I(u_{tj}(\theta_0) \leq v_j) - G(v|\mathcal{Z}_0^{t-1}, \theta_0)], \tag{3.14}$$

$$q_{2t}(\theta, \gamma) = w(\sum_{j=0}^{t} \xi_j \phi(z_{t-j}))[\sum_{j=1}^{r} c_j I(u_{tj}(\theta_0) \leq v_j) - G(v|\mathcal{Z}_0^{t-1}, \theta)], \tag{3.15}$$

$$q_{3t}(\theta, \gamma) = w(\sum_{j=0}^{t} \xi_j \phi(z_{t-j}))[\sum_{j=1}^{r} c_j I(u_{tj}(\theta) \leq v_j) - G(v|\mathcal{Z}_0^{t-1}, \theta_0)], \tag{3.16}$$

$$q_{4t}(\theta, \gamma) = w(\sum_{j=0}^{t} \xi_j \phi(z_{t-j}))[\sum_{j=1}^{r} c_j I(u_{tj}(\theta) \leq v_j) - G(v|\mathcal{Z}_0^{t-1}, \theta)]. \tag{3.17}$$

We consider the Cramér-von Mises statistics $S_{iT} = \int_{\gamma\in\Gamma} Q_{iT}(\gamma)^2 d\mu(\gamma)$, where $Q_{1T}(\gamma) = T^{-\frac{1}{2}}\sum_{t=1}^{T} q_{1t}(\gamma)$, $Q_{iT}(\gamma) = T^{-\frac{1}{2}}\sum_{t=1}^{T} q_{it}(\hat{\theta}_T,\gamma)$ for $i = 2,3,4$, and $\mu(\cdot)$ is a probability measure on , which is absolutely continuous with respect to the Lebesgue measure. [2] When G and g do not depend on θ (no parameter uncertainty), S_{1T} should be used. When only G depends on θ, S_{2T} should be used. When only g depends on θ, S_{3T} should be used. When both G and g depend on θ, S_{4T} should be used.

THEOREM 3.1: Suppose that Assumption 1 holds.

(a) Under the null hypothesis (2.1),

$$S_{iT} \xrightarrow{d} S_i \equiv \int_{\gamma\in\Gamma} Q_i(\gamma)^2 d\mu(\gamma)$$

where Q_i is a mean-zero Gaussian process with covariance kernel:

$$E[Q_i(\gamma)Q_i(\gamma')] = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E[r_t(\gamma)r_t(\gamma')] \quad (3.18)$$

for $i = 1,2,3,4$, where

$$\begin{aligned}
r_{1t}(\gamma) &= q_{1t}(\gamma)\\
r_{2t}(\gamma) &= q_{1t}(\gamma) - E[w(\sum_{j=1}^{\infty}\xi_j\phi(z_{-j}))G_\theta(v|\mathcal{Z}_{-\infty}^{-1},\theta_0)]\psi_t(z_t,...,z_0,\theta_0)\\
r_{3t}(\gamma) &= q_{1t}(\gamma) + E[w(\sum_{j=1}^{\infty}\xi_j\phi(z_{-j}))G_v(v|\mathcal{Z}_{-\infty}^{-1},\theta_0)g_\theta(z_0,\theta_0)]\psi_t(z_t,...,z_0,\theta_0)\\
r_{4t}(\gamma) &= q_{1t}(\gamma) + E[w(\sum_{j=1}^{\infty}\xi_j\phi(z_{-j}))(G_v(v|\mathcal{Z}_{-\infty}^{-1},\theta_0)g_\theta(z_0,\theta_0) - G_\theta(v|\mathcal{Z}_{-\infty}^{-1},\theta_0))]\psi_t(z_t,...,z_0,\theta_0)
\end{aligned}$$

(b) Under the alternative hypothesis (2.2),

$$\text{plim}_{T\to\infty}\frac{1}{T}\int_{\gamma\in\Gamma} Q_{iT}(\gamma)^2 d\mu(\gamma) > 0$$

[2]When Υ is a singleton (e.g., interval forecast restrictions), the average-type test statistic is defined as $\int_{\xi\in\Xi} Q_{iT}(\gamma)^2 d\xi$.

for $i = 1, 2, 3, 4$.

Theorem 3.1(a) shows that the null limiting distribution is data-dependent, and Theorem 3.1(b) shows that the test statistic diverges under the alternative hypothesis. In order to make the test operational, we follow the approach of Bierens and Ploberger (1997). By Lemma 1 and Theorem 7 in Bierens and Ploberger (1997), we can find the upperbounds as follows.[3]

$$\lim_{T\to\infty} P\left(\int Q_i(\gamma)/V_i(\gamma)d\mu(\gamma) > \delta\right) \leq P\left(\sup_{n\geq 1}\frac{1}{n}\sum_{j=1}^{n}\epsilon_j^2 > \delta\right) \tag{3.19}$$

where $V_i(\gamma) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} Var(r_{it}(\gamma))$ and $\epsilon_j \sim NID(0,1)$. Thus, we need a consistent estimator for $V_i(\gamma)$. Let

$$V_{1T}(\gamma) = \frac{1}{T}\sum_{t=1}^{T} w(\sum_{j=0}^{t}\xi_j\phi(z_{t-j}))^2[\sum_{j=1}^{r} c_j I(u_{tj}(\theta_0)\leq v_j) - G(v|\mathcal{Z}_0^{t-1},\theta_0)]^2$$

$$V_{2T}(\gamma) = \frac{1}{T}\sum_{t=1}^{T}\Big(w(\sum_{j=0}^{t}\xi_j\phi(z_{t-j}))[\sum_{j=1}^{r} c_j I(u_{tj}(\theta_0)\leq v_j) - G(v|\mathcal{Z}_0^{t-1},\hat{\theta}_T)]$$
$$-\frac{1}{T}\sum_{s=1}^{T}\{w(\sum_{j=0}^{s}\xi_j\phi(z_{s-j}))G_\theta(v|\mathcal{Z}_0^{s-1},\hat{\theta}_T)]\}\psi(z_t,...,z_0,\hat{\theta}_T)\Big)^2,$$

$$V_{3T}(\gamma) = \frac{1}{T}\sum_{t=1}^{T}\Big(w(\sum_{j=0}^{t}\xi_j\phi(z_{t-j}))[\sum_{j=1}^{r} c_j I(u_{tj}(\theta_0)\leq v_j) - G(v|\mathcal{Z}_0^{t-1},\hat{\theta}_T)]$$
$$+\frac{1}{T}\sum_{s=1}^{T}\{w(\sum_{j=0}^{s}\xi_j\phi(z_{s-j}))G_v(v|\mathcal{Z}_0^{s-1},\hat{\theta}_T)]\}\psi(z_t,...,z_0,\hat{\theta}_T)\Big)^2,$$

$$V_{4T}(\gamma) = \frac{1}{T}\sum_{t=1}^{T}\Big(w(\sum_{j=0}^{t}\xi_j\phi(z_{t-j}))[\sum_{j=1}^{r} c_j I(u_{tj}(\theta_0)\leq v_j) - G(v|\mathcal{Z}_0^{t-1},\hat{\theta}_T)]$$
$$+\frac{1}{T}\sum_{s=1}^{T}\{w(\sum_{j=0}^{s}\xi_j\phi(z_{s-j}))(G_v(v|\mathcal{Z}_0^{s-1},\hat{\theta}_T) - G_\theta(v|\mathcal{Z}_0^{s-1},\hat{\theta}_T))]\}\psi(z_t,...,z_0,\hat{\theta}_T)\Big)^2,$$

$$V_{5T}(\gamma) = \frac{1}{T}\sum_{t=1}^{T}\Big(w(\sum_{j=0}^{t}\xi_j\phi(z_{t-j}))[\sum_{j=1}^{r} c_j I(u_{tj}(\hat{\theta}_T)\leq v_j) - G(v|\mathcal{Z}_0^{t-1},\theta_0)]$$

[3] When , is not compact, use the change-of-variable technique so that the new integration is taken over a compact support.

$$+\frac{1}{2h_T T}\sum_{s=1}^{T} w(\sum_{j=0}^{s}\xi_j\phi(z_{s-j}))[\hat{G}_{\upsilon s}(\hat{\theta}_T)g_\theta(z_s,\hat{\theta}_T)\psi(z_t,...,z_0,\hat{\theta}_T)\Big)^2,$$

where

$$\hat{G}_{\upsilon t}(\theta)=(\hat{G}_{\upsilon t}^{(1)}(\theta),\hat{G}_{\upsilon t}^{(2)}(\theta),...,\hat{G}_{\upsilon t}^{(r)}(\theta))',\ \hat{G}_{\upsilon t}^{(i)}(\theta)=c_i I(|(u_{ti}(\theta)-\upsilon_i)/h_T|\leq 1).$$

THEOREM 3.2: Suppose that Assumption 1 holds.

(a) Under the null hypothesis (2.1), $\sup_{\gamma\in\Gamma}|V_{iT}(\gamma)-V_i(\gamma)| = o_p(1)$ for $i=1,2,3,4$. Under the alternative hypothesis (2.2), $\sup_{\gamma\in\Gamma}|V_{iT}(\gamma)| = O_p(1)$ for $i=1,2,3,4$.

(b) Moreover, assume that $h_T\to 0$ and $h_T T^{1/2}\to\infty$ as $T\to\infty$. Under the null hypothesis (2.1), $\sup_{\gamma\in\Gamma}|V_{5T}(\gamma)-V_3(\gamma)| = o_p(1)$. Under the alternative hypothesis (2.2), $\sup_{\gamma\in\Gamma}|V_{5T}(\gamma)| = O_p(1)$.

$V_{5T}(\cdot)$ should be used when the null hypothesis is not parametric.

4. Monte Carlo Experiments

This section investigates the finite-sample performance of the proposed test for conditional goodness-of-fit restrictions in a simple set of Monte Carlo experiments. The null hypothesis is

$$y_t \sim N(\mu,\sigma^2) \tag{4.20}$$

for some μ and $sigma^2>0$. In the notation used in Section 3, $u_t(\theta_0)=(y_t-\mu_0)/\sigma$ and $z_t=y_t$.

Several alternative hypotheses are considered. Then the first alternative hypothesis is

$$y_t=\varepsilon_t^2-1 \tag{4.21}$$

where $\varepsilon_t \sim NID(0,1)$. The second is a stochastic volatility model, i.e.,

$$y_t = \exp(h_t/2)\varepsilon_t,\ h_t = c + \rho h_{t-1} + \sigma\eta_t, \tag{4.22}$$

where $(\varepsilon_t, \eta_t)^\top \sim NID(\mathbf{0}_2, I_2)$. The third is an AR(1) model, i.e.,

$$y_t = \beta y_{t-1} + \varepsilon_t \tag{4.23}$$

where $\varepsilon_t \sim N(0,1)$.

Without loss of generality, $\mu_0 = 0$ and $\sigma^2 = 1$. The sample sizes used are $T = 250, 500, 1000$. We use $\Upsilon = \Re$ and $\Xi_j = [-j^{-2}, j^{-2}]$ for $j = 1, 2, ..., T$. The Monte Carlo integration method is used to compute the Cramér-von Mises statistic. 300 random vectors, consisting of a standard normal r.v. and T uniform r.v.s over $\Upsilon \times \Xi$, are drawn for each Monte Carlo replication. The number of Monte Carlo replications is set to 1000.

Table 1 shows the rejection frequencies (%). The size here is a pointwise one (for fixed μ and σ^2) and thus is different from the size in the sense of Lehmann (1986): the supremum of the type I error probabilities over all possible DGP under the null (for all possible values of μ and σ^2). Because we are using the upperbound, it is not surprising that the actual rejection frequenciese are lower than the nominal level. In fact, the actual rejection frequencies turn out to be zero. The test is not powerful when the sample size is small. However, the power of the test increases as the sample size increases. The test is practically powerful enough for the sample size 1000. We also note that the further the alternative hypothesis is from the null, the more powerful the test is. In practice, the use of the 10% significance level is recommended to increase the power. As Lehmann (1986), one may want to use higher values of the significant level than the customary ones, such as 10%.

Appendix: Proofs of Theorems 3.1 and 3.2

We use the following notation in the proof. Let $w_{\max}$ and $w_{\min}$ denote constants such that

$$\sup_{\xi\in\Xi} |w(\sum_{j=1}^{\infty} \xi_j \phi(z_{-j}))| \le w_{\max} \text{ a.s. and } \inf_{\xi\in\Xi} |w(\sum_{j=1}^{\infty} \xi_j \phi(z_{-j}))| \ge w_{\min} \text{ a.s.},$$

respectively. Let $\tilde{w}$ denote a constant such that $|w(x') - w(x)| \le \tilde{w}|x' - x|$ for all $x, x' \in [\kappa_{\min}, \kappa_{\max}]$ where $-\infty < \kappa_{\min} \le \inf_{\xi\in\Xi} \sum_{j=1}^{\infty} \xi_j \phi(z_{-j})$ a.s. and $\sup_{\xi\in\Xi} \sum_{j=1}^{\infty} \xi_j \phi(z_{-j}) \le \kappa_{\max} < \infty$ a.s. Assumptions 1(e) and (f) guarantee the existence of $w_{\max}$, $w_{\min}$, $\tilde{w}$, $\kappa_{\max}$, and $\kappa_{\min}$. $\mathbf{0}_d$ and $\mathbf{1}_d$ denote d-dimensional vectors such that $\mathbf{0}_d = (0, 0, ..., 0)^{\top}$ and $\mathbf{1}_d = (1, 1, ..., 1)^{\top}$, respectively. When applied to vectors, max and min are element-by-element. To simplify notation, we assume that θ is scalar, i.e., $p = 1$, and $\Xi_j = [\xi_{\min j}, \xi_{\max j}]^d$ for some $\xi_{\min j}$¡$\xi_{\max j}$ for $j = 1, 2, ..., r$. **Unless noted otherwise, O_{as}, o_{as}, O_p, and o_p are uniform in** $\gamma \in$, **(e.g., $f(\gamma) = o_p(1)$ means $\sup_{\gamma\in\Gamma} |f(\gamma)| = o_p(1)$).**

Proof of Theorem 3.1(a): Without loss of generality, we prove only the case when $i = 4$. For notational simplicity, we omit the subscript 4. The proof of Theorem 3.1(a) consists of the following five lemmas.

LEMMA A.1: (, ρ) is totally bounded, where ρ is a metric on , defined by

$$\rho(\gamma, \gamma') = \Big(\sum_{j=1}^{\infty} |\xi_j' - \xi_j|^2 j^{1+\zeta} + \sum_{j=1}^{r} |F_j(v_j') - F_j(v_j)|^2 \Big)^{\frac{1}{2}}.$$

where F_j is the marginal d.f. of $u_{tj}(\theta_0) = g_j(z_t, \theta_0)$.

LEMMA A.2:

$$\frac{1}{T}\sum_{t=1}^{T} w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j})) G_v(v|\mathcal{Z}_0^{t-1}, \bar{\theta}_T) g_\theta(z_t, \bar{\theta}_T) - E[w(\sum_{j=1}^{\infty} \xi_j \phi(z_{-j})) G_v(v|\mathcal{Z}_{-\infty}^{-1}, \theta_0) g_\theta(z_0, \theta_0)] = o_p(1), \tag{A.1}$$

$$\frac{1}{T}\sum_{t=1}^{T} w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j})) G_\theta(v|\mathcal{Z}_0^{t-1}, \bar{\theta}_T) - E[w(\sum_{j=1}^{\infty} \xi_j \phi(z_{-j})) G_\theta(v|\mathcal{Z}_{-\infty}^{-1}, \theta_0)] = o_p(1), \tag{A.2}$$

where $\{\bar{\theta}_T\}_{T=1}^{\infty}$ is an arbitrary sequence such that $\bar{\theta}_T - \theta_0 = o_p(1)$ uniformly in $v \in \Upsilon$. Moreover,

$$E\Big\{ w(\sum_{j=1}^{\infty} \xi_j \phi(z_{-j}))[G_v(v|\mathcal{Z}_{-\infty}^{-1}, \theta_0) g_\theta(z_0, \theta_0) + G_\theta(v|\mathcal{Z}_{-\infty}^{-1}, \theta_0)] \Big\} \tag{A.3}$$

is uniformly continuous in $\gamma \in$, .

LEMMA A.3: $\{T^{-\frac{1}{2}} \sum_{t=1}^{T} q_t(\theta_0, \gamma) : \gamma \in$, $\}_{t=1}^{T}$ is stochastically equicontinuous, i.e., for each $\epsilon > 0$,

$$\lim_{\delta\downarrow 0} \limsup_{T\to\infty} P\left(\sup_{\gamma,\gamma'\in\Gamma:\rho(\gamma,\gamma')<\delta} \left| T^{-\frac{1}{2}} \sum_{t=1}^{T} [q_t(\theta_0, \gamma') - q_t(\theta_0, \gamma)] \right| > \epsilon \right) = 0. \tag{A.4}$$

LEMMA A.4:

$$T^{-\frac{1}{2}} \sum_{t=1}^{T} \Big\{ q_t(\hat{\theta}_T, \gamma) - r_t(\gamma) \Big\} = o_p(1). \tag{A.5}$$

LEMMA A.5:

$$T^{-\frac{1}{2}} \sum_{t=1}^{T} r_t(\gamma) \xrightarrow{d} N(0, V(\gamma)) \tag{A.6}$$

for every $\gamma \in$, , where $V(\gamma) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} Var(r_t(\gamma))$.

Before we prove these lemmas, we shall briefly sketch the proof of Theorem 3.1(a). By Lemma A.4, $T^{-1/2}\sum_{t=1}^{T} r_t(\gamma)$ approximates $T^{-1/2}\sum_{t=1}^{T} q_t(\hat{\theta}_T, \gamma)$ uniformly in $\gamma \in$, . By Lemma A.1, $(, , \rho)$ is totally bounded. Generalizing Lemma A.5 and applying the Cramér-Wold device, one can show the fi-di convergence of $T^{-1/2}\sum_{t=1}^{T} r_t(\gamma)$ to $Q(\gamma)$. By Lemmas A.2 and A.3, $T^{-1/2}\sum_{t=1}^{T} r_t(\gamma)$ is stochastically equicontinuous. Therefore, $\{Q_T(\gamma) : \gamma \in , \}_{t=1}^{\infty}$ weakly converges to $\{Q(\gamma) : \gamma \in , \}$.

Proof of Lemma A.1: The proof is a slight generalization of and analogous to that of Lemma 1 in De Jong (1996) and thus is omitted.

Proof of Lemma A.2: Since the proof of (A.2) is analogous to that of (A.1), we prove only (A.1). It suffices to show

$$\frac{1}{T}\sum_{t=1}^{T}[w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))G_v(v|\mathcal{Z}_0^{t-1},\bar{\theta}_T)g_\theta(z_t,\bar{\theta}_T) - w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))G_v(v|\mathcal{Z}_0^{t-1},\theta_0)g_\theta(z_t,\theta_0)] = o_p(1), \tag{A.7}$$

$$\frac{1}{T}\sum_{t=1}^{T} w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))G_v(v|\mathcal{Z}_0^{t-1},\theta_0)g_\theta(z_t,\theta_0) - E[w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))G_v(v|\mathcal{Z}_0^{t-1},\theta_0)g_\theta(z_t,\theta_0)] = o_{as}(1), \tag{A.8}$$

and

$$\frac{1}{T}\sum_{t=1}^{T} E[w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))G_v(v|\mathcal{Z}_0^{t-1},\theta_0)g_\theta(z_t,\theta_0)] - E[w(\sum_{j=1}^{\infty}\xi_j\phi(z_{-j}))G_v(v|\mathcal{Z}_{-\infty}^{-1},\theta_0)g_\theta(z_0,\theta_0)] = o(1). \tag{A.9}$$

First, we shall prove (A.7).

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\left|\frac{1}{T}\sum_{t=1}^{T}[w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))G_v(v|\mathcal{Z}_0^{t-1},\bar{\theta}_T)g_\theta(z_t,\bar{\theta}_T) - w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))G_v(v|\mathcal{Z}_0^{t-1},\theta_0)g_\theta(z_t,\theta_0)]\right| \\
\leq\ & \frac{2w_{\max}}{T}\sum_{t=1}^{T}\bar{G}_t(z_{t-1},...,z_0)\bar{g}(z_t)|\bar{\theta}_T-\theta_0| = o_p(1).
\end{aligned} \tag{A.10}$$

The inequality follows from Assumptions 1(b)(c), the triangle inequality, and definition of $w_{\max}$. The equality follows from Assumptions 1(a)(b)(c)(d) and Theorem 3.1 in McLeish (1975).

Second, we shall prove (A.8). By the uniform SLLN, it suffices to show the total boundedness of $(, , \rho)$, the stochastic equicontinuity

$$\begin{aligned}
\lim_{\delta\downarrow 0}\limsup_{T\to\infty}\sup_{\gamma,\gamma'\in\Gamma:\rho(\gamma,\gamma')<\delta}&\left|\frac{1}{T}\sum_{t=1}^{T}[w(\sum_{j=1}^{t}\xi_j'\phi(z_{t-j}))G_v(v'|\mathcal{Z}_0^{t-1},\theta_0)g_\theta(z_t,\theta_0)\right. \\
&\left.-w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))G_v(v|\mathcal{Z}_0^{t-1},\theta_0)g_\theta(z_t,\theta_0)]\right| = 0 \text{ a.s.}
\end{aligned} \tag{A.11}$$

and the pointwise almost sure convergence

$$\frac{1}{T}\sum_{t=1}^{T}\left\{w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))G_v(v|\mathcal{Z}_0^{t-1},\theta_0)g_\theta(z_t,\theta_0) - E[w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))G_v(v|\mathcal{Z}_0^{t-1},\theta_0)g_\theta(z_t,\theta_0)]\right\} = o_{as}(1), \tag{A.12}$$

for every $\gamma \in$, . (, , ρ) is totally bounded by Lemma A.1. It is straightforward to show the stochastic equicontinuity (A.11) by using Assumption 1(a)(b)(e)(f), Theorem 3.1 in McLeish (1975) and Lemma A.1; therefore, the proof is omitted. We shall apply Theorem 3.1 in McLeish (1975) in order to show (A.12). Thus, it suffices to show that

$$\left\{ w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j})) G_v(v|\mathcal{Z}_0^{t-1}, \theta_0) g_\theta(z_t, \theta_0) \right\}_{t=1}^{\infty} \tag{A.13}$$

is L^2-NED on $\{z_t\}_{t=1}^{\infty}$ of size $-1/2$.

$$\begin{aligned}
& \sup_t \left\| w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j})) G_v(v|\mathcal{Z}_0^{t-1}, \theta_0) g_\theta(z_t, \theta_0) - E[w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j})) G_v(v|\mathcal{Z}_0^{t-1}, \theta_0) g_\theta(z_t, \theta_0) | \mathcal{Z}_{t-m}^{t+m}] \right\|_2 \\
= \; & \sup_t \left\| w(\sum_{j=1}^{t} \xi_j \phi(z_{-j})) G_v(v|\mathcal{Z}_{-t}^{-1}, \theta_0) g_\theta(z_0, \theta_0) - E[w(\sum_{j=1}^{t} \xi_j \phi(z_{-j})) G_v(v|\mathcal{Z}_{-t}^{-1}, \theta_0) g_\theta(z_0, \theta_0) | \mathcal{Z}_{-m}^{m}] \right\|_2 \\
\leq \; & \sup_t \left\| w(\sum_{j=1}^{m} \xi_j \phi(z_{-j}))(G_v(v|\mathcal{Z}_{-t}^{-1}, \theta_0) g_\theta(z_0, \theta_0) - E[G_v(v|\mathcal{Z}_{-t}^{-1}, \theta_0) g_\theta(z_0, \theta_0) | \mathcal{Z}_{-m}^{m}]) \right\|_2 \\
& + \sup_t \left\| \tilde{w} \sum_{j=m+1}^{t} |\xi_j| \{\phi(z_{-j}) G_v(v|\mathcal{Z}_{-t}^{-1}, \theta_0) g_\theta(z_0, \theta_0) - E[\phi(z_{-j}) G_v(v|\mathcal{Z}_{-t}^{-1}, \theta_0) g_\theta(z_0, \theta_0)]\} \right\|_2 . \\
\leq \; & w_{\max} \sup_t \left\| G_v(v|\mathcal{Z}_{-t}^{-1}, \theta_0) g_\theta(z_0, \theta_0) - E[G_v(v|\mathcal{Z}_{-t}^{-1}, \theta_0) g_\theta(z_0, \theta_0) | \mathcal{Z}_{-m}^{m}] \right\|_2 && \text{(A.14)} \\
& + \tilde{w} \sum_{j=m+1}^{\infty} |\xi_j| \sup_t \left\| \phi(z_{-j})[G_v(v|\mathcal{Z}_{-t}^{-1}, \theta_0) g_\theta(z_0, \theta_0) - E[\phi(z_{-j}) G_v(v|\mathcal{Z}_{-t}^{-1}, \theta_0) g_\theta(z_0, \theta_0)] \right\|_2 . && \text{(A.15)}
\end{aligned}$$

The equality follows from the strict stationarity, the first inequality from the triangle inequality, and the second inequatlity from the Cauchy-Schwarz inequatlity. (A.14) is of size $-1/2$ by Assumption 1(b). By Assumption 1(a)(b)(c)(e)(f), $\sup_t \| \cdot \|_2$ in (A.15) is finite uniformly in j and t. Because

$$\sup_{\xi \in \Xi} \sum_{j=m+1}^{\infty} |\xi_j| = C \sum_{j=m+1}^{\infty} j^{-1-\zeta} = O(m^{-\zeta}) \tag{A.16}$$

for some $C > 0$ by Assumption 1(e), (A.15) is of size $-1/2$. Thus, (A.13) is L^2-NED on $\{z_t\}_{t=1}^{\infty}$ of size $-1/2$. Therefore, the uniform SLLN proves (A.8).

Third, we shall prove (A.9).

$$\begin{aligned}
& \left| \frac{1}{T} \sum_{t=1}^{T} \{E[w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j})) G_v(v|\mathcal{Z}_0^{t-1}, \theta_0) g_\theta(z_t, \theta_0)] - E[w(\sum_{j=1}^{\infty} \xi_j \phi(z_{-j})) G_v(v|\mathcal{Z}_{-\infty}^{-1}, \theta_0) g_\theta(z_0, \theta_0)]\} \right| \\
= \; & \left| \frac{1}{T} \sum_{t=1}^{T} \{E[w(\sum_{j=1}^{t} \xi_j \phi(z_{-j})) G_v(v|\mathcal{Z}_{-t}^{-1}, \theta_0) g_\theta(z_0, \theta_0)] - E[w(\sum_{j=1}^{\infty} \xi_j \phi(z_{-j})) G_v(v|\mathcal{Z}_{-t}^{-1}, \theta_0) g_\theta(z_0, \theta_0)]\} \right| \\
\leq \; & \left| \frac{1}{T} \sum_{t=1}^{T} E[w(\sum_{j=1}^{t} \xi_j \phi(z_{-j})) - w(\sum_{j=1}^{\infty} \xi_j \phi(z_{-j}))] G_v(v|\mathcal{Z}_{-t}^{-1}, \theta_0) g_\theta(z_0, \theta_0) \right| \\
& + \left| \frac{1}{T} \sum_{t=1}^{T} E w(\sum_{j=1}^{\infty} \xi_j \phi(z_{-j}))[G_v(v|\mathcal{Z}_{-t}^{-1}, \theta_0) - G_v(v|\mathcal{Z}_{-\infty}^{-1}, \theta_0)] g_\theta(z_0, \theta_0)] \right|
\end{aligned}$$

$$\leq \quad \Big[\frac{\tilde{w}^2}{T} \sum_{t=1}^{T} \big(\sum_{j=t+1}^{\infty} |\xi_j|^2 j^{1+\zeta} \big) \big(\sum_{j=t+1}^{\infty} E|\phi(z_{-j})|^2 j^{-1-\zeta} \big) \Big]^{\frac{1}{2}} \Big\{ \frac{1}{T} \sum_{t=1}^{T} E[\bar{G}_t(z_{-1},...,z_{-t})\bar{g}(z_0)]^2 \Big\}^{\frac{1}{2}} \quad (A.17)$$

$$+\frac{w_{\max}\|\bar{g}_\theta(z_0)\|_2}{T} \sum_{t=1}^{T} \left\| G_v(v|\mathcal{Z}_{-t}^{-1},\theta_0) - G_v(v|\mathcal{Z}_{-\infty}^{-1},\theta_0) \right\|_2 . \quad (A.18)$$

The equality follows from strict stationarity, the first inequality from the triangle inequality, and the second inequality from the Cauchy Schwarz inequality. (A.17) goes to zero uniformly in $\gamma \in$,

$$\lim_{T\to\infty} \sum_{t=1}^{T} t^{-1} \sum_{j=t+1}^{\infty} |\xi_j|^2 j^{1+\zeta} \sum_{j=t+1}^{\infty} E|\phi(z_{-j})|^2 j^{-1-\zeta} < \infty. \quad (A.19)$$

(A.18) goes to zero uniformly in $\gamma \in$, by Assumption 1(b). The proof of the uniform continuity of (A.3) is straightforward and thus is omitted.

Proof of Lemma A.3: As in the literature on empirical d.f.s (e.g., Bai, 1996, p.610), we focus on the case $\Upsilon = [0,1]^r$ in terms of stochastic equicontinuity. For an arbitrary $\delta > 0$ and $T \in N$, let $(\Delta_T(\delta), J_T, \{K_{Tj}\}_{j=1}^{J_T}, L_T)$ be such that $\sup_{\xi,\xi'\in\Xi} \sum_{j=J_T+1}^{\infty} |\xi_j - \xi| \leq \Delta_T(\delta)$, $L_T = [1/\Delta_T(j)]$, $K_{Tj} = [(\xi_{\max j} - \xi_{\min j})/\Delta_T(\delta)]$, $\Delta_T(\delta) = \delta/\max\{(\Pi_{j=1}^{J_T} K_{Tj}^{2d}) L_T^{2r}, T^{1/2}\}$. One can show that, as $T \to \infty$, $\Delta_T(\delta) \to 0$, $J_T \to \infty$, $K_{Tj} \to \infty$, and $L_T \to \infty$. For $j = 1, 2, ..., J_T$, let

$$\xi_j(\mathbf{k}) = \min(\xi_{\min j}\mathbf{1}_d + \Delta_T(\delta)\mathbf{k}, \xi_{\max j})\mathbf{1}_d, \text{ for } \mathbf{0}_d \leq \mathbf{k} \leq \mathbf{K}_{Tj}.$$

For all $j > J_T$, let $\xi_j(\mathbf{k}) = \xi_j^*$ for some $\xi_j^* \in \Xi_j$. For $\mathbf{0}_r \leq \mathbf{l} \leq \mathbf{L}_T$, let

$$v(\mathbf{l}) = \min\{\mathbf{0}_r + \Delta_T(\delta)\mathbf{l}, \mathbf{1}_r\}.$$

Because

$$\begin{aligned}
&\sup_{\gamma,\gamma'\in\Gamma:\rho(\gamma,\gamma')<\delta} \left| T^{-\frac{1}{2}} \sum_{t=1}^{T} [q_t(\theta_0,\gamma') - q_t(\theta_0,\gamma)] \right| \\
\leq \quad & 3 \max_{1\leq k,k'\leq K_T, 1\leq l,l'\leq L_T:\rho(\gamma(k,l),\gamma(k',l'))<\delta} \left| T^{-\frac{1}{2}} \sum_{t=1}^{T} [q_t(\theta_0,\gamma(\mathbf{k},\mathbf{l})) - q_t(\theta_0,\gamma(\mathbf{k},\mathbf{l})))] \right| \\
&+2w_{\max}T^{-1} \sum_{t=1}^{T} \bar{G}_t(z_{t-1}, z_{t-2}, ...z_0)\delta \\
&+2|c|\tilde{w}T^{-1} \sum_{t=1}^{T} \sum_{j=1}^{\infty} j^{-1-\zeta} |\phi(z_{t-j})|\delta. \quad (A.20)
\end{aligned}$$

where the inequality follows from the triangle inequality, the Cauchy-Schwarz inequality, Assumption 1(b), and the definition of $\tilde{w}$,

$$\begin{aligned}
& P\left(\sup_{\gamma,\gamma'\in\Gamma:\rho(\gamma,\gamma')<\delta} \left| T^{-\frac{1}{2}} \sum_{t=1}^{T} [q_t(\theta_0,\gamma') - q_t(\theta_0,\gamma)] \right| < \varepsilon \right) \\
\leq \quad & P\left(3 \max_{k,k',l,l':\rho(\gamma(k,l),\gamma(k',l'))<\delta} \left| T^{-\frac{1}{2}} \sum_{t=1}^{T} [q_t(\theta_0,\gamma(\mathbf{k'},\mathbf{l'})) - q_t(\theta_0,\gamma(\mathbf{k},\mathbf{l})))] \right| < \varepsilon/3 \right) \quad (A.21) \\
& +P\left(2w_{\max}T^{-1} \sum_{t=1}^{T} |\bar{G}_t(z_{t-1}, z_{t-2}, ...z_0)|\delta < \varepsilon/3 \right) \quad (A.22)
\end{aligned}$$

$$+P\left(2|c|\tilde{w}T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{\infty}j^{-1-\zeta}|\phi(z_{t-j})|\delta<\varepsilon/3\right). \tag{A.23}$$

Since the limsups of (A.22) and (A.23) are $O(\delta)$, it suffices to show that (A.21) is $O(\delta)$. It follows from

$$\begin{aligned}
& P\left(3\max_{k,k',l,l':\rho(\gamma(k,l),\gamma(k',l'))<\delta}\left|T^{-\frac{1}{2}}\sum_{t=1}^{T}[q_t(\theta_0,\gamma(\mathbf{k'},\mathbf{l'}))-q_t(\theta_0,\gamma(\mathbf{k},\mathbf{l})))]\right|<\varepsilon/3\right) \\
\leq\ & K_T^2L_T^2\max_{k,k',l,l'}P\left(\left|T^{-\frac{1}{2}}\sum_{t=1}^{T}[q_t(\theta_0,\gamma(\mathbf{k'},\mathbf{l'}))-q_t(\theta_0,\gamma(\mathbf{k},\mathbf{l})))]\right|<\varepsilon/3\right) \\
\leq\ & K_T^2L_T^2\max_{k,k',l,l'}(\varepsilon/3)^{-2}T^{-1}\sum_{t=1}^{T}E[q_t(\theta_0,\gamma(\mathbf{k'},\mathbf{l'}))-q_t(\theta_0,\gamma(\mathbf{k},\mathbf{l})))]^2 \qquad \text{(A.24)}
\end{aligned}$$

and

$$E[q_t(\theta_0,\gamma(\mathbf{k},\mathbf{l}))-q_t(\theta_0,\gamma(\mathbf{k},\mathbf{l})))]^2\leq 2w_{\max}^2\bar{G}_t(z_{t-1},...,z_0)\Delta_T(\delta)+2|c|w_{\max}\tilde{w}\Delta_T(\delta) \tag{A.25}$$

that

$$\begin{aligned}
& P\left(3\max_{k,k',l,l':\rho(\gamma(k,l),\gamma(k',l'))<\delta}\left|T^{-\frac{1}{2}}\sum_{t=1}^{T}[q_t(\theta_0,\gamma(\mathbf{k'},\mathbf{l'}))-q_t(\theta_0,\gamma(\mathbf{k},\mathbf{l})))]\right|<\varepsilon/3\right) \\
\leq\ & 2\varepsilon^{-2}T^{-1}\sum_{t=1}^{T}[\bar{G}_t(z_{t-1},...,z_0)+|c|w_{\max}\tilde{w}]\delta. \qquad \text{(A.26)}
\end{aligned}$$

Therefore, Lemma A.3 is proved. *Q.E.D.*

Proof of Lemma A.4: By Lemma A.2, it suffices to show

$$T^{-\frac{1}{2}}\sum_{t=1}^{T}w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))\{\sum_{j=1}^{r}c_j[I(u_{tj}(\hat{\theta}_T)\leq v_j)-I(u_{tj}(\theta_0)\leq v_j)]-G(\nu_t|\mathcal{Z}_0^{t-1},\theta_0)+G(\nu|\mathcal{Z}_0^{t-1},\theta_0)\}=o_p(1) \tag{A.27}$$

$$T^{-\frac{1}{2}}\sum_{t=1}^{T}w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))[G(\nu_t|\mathcal{Z}_0^{t-1},\theta_0)-G(\nu|\mathcal{Z}_0^{t-1},\theta_0)-G_v(v|\mathcal{Z}_{-\infty}^{t-1},\bar{\theta}_T)g_\theta(z_t,\bar{\theta}_T)(\hat{\theta}_T-\theta_0)]=o_p(1), \tag{A.28}$$

and

$$T^{-\frac{1}{2}}\sum_{t=1}^{T}w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))[G(v|\mathcal{Z}_0^{t-1},\hat{\theta}_T)-G(v|\mathcal{Z}_0^{t-1},\theta_0)-G_\theta(v|\mathcal{Z}_0^{t-1},\tilde{\theta}_T)(\hat{\theta}_T-\theta_0)]=o_p(1), \tag{A.29}$$

for some sequences $\{\bar{\theta}_T\}_{T=1}^{\infty}$ and $\{\tilde{\theta}_T\}_{T=1}^{\infty}$ such that $\bar{\theta}_T-\theta_0=o_{as}(1)$ and $\tilde{\theta}_T-\theta_0=o_{as}(1)$ uniformly in $v\in\Upsilon$, respectively. Let $v_t=v+g(z_t,\hat{\theta}_T)-g(z_t,\theta_0)$. (A.27) holds by the stochastic equicontinuity (Lemma A.3). (A.28) and (A.29) follow from the mean-value theorem. *Q.E.D.*

Proof of Lemma A.5: Exploiting the fact that $\{(\sum_{t=1}^{T}r_t(\gamma),\mathcal{Z}_0^{T-1})\}_{T=1}^{\infty}$ is a martingale, we shall apply the CLT for martingales in Hall and Heyde (1980, Theorem 3.2). We need to show

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}[r_t(\gamma)^2-Er_t(\gamma)^2]=0. \tag{A.30}$$

By Theorem 3.1 in McLeish (1988), it suffices to show that $\{r_t(\gamma)^2\}_{t=1}^{\infty}$ is L^2-NED on $\{z_t\}_{t=-\infty}^{\infty}$ of size $-1/2$. Since

$$\{E[w(\sum_{j=1}^{\infty}\xi_j\phi(z_{-j}))G_v(v|\mathcal{Z}_{-\infty}^{-1},\theta_0)g_\theta(z_0,\theta_0)+G_\theta(v|\mathcal{Z}_{-\infty}^{-1},\theta_0)]\psi_t(z_t,z_{t-1},...,z_0,\theta_0)\}_{t=1}^{\infty}$$

is L^2-NED of size $-1/2$ on $\{z_t\}$, it suffices to show that $\{q_t(\theta_0,\gamma)^2\}_{t=1}^{\infty}$ is L^2-NED on $\{z_t\}_{t=-\infty}^{\infty}$ of size $-1/2$.

$$\begin{aligned}
& \sup_t \Big\| \, w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))^2 [\sum_{j=1}^{r} c_j I(u_{tj}(\theta_0) \le v_j) - G(v|\mathcal{Z}_0^{t-1},\theta_0)]^2 \\
& -E\{w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))^2 [\sum_{j=1}^{r} c_j I(u_{tj}(\theta_0) \le v_j) - G(v|\mathcal{Z}_0^{t-1},\theta_0)]^2 |\mathcal{Z}_{t-m}^{t+m}\} \, \Big\|_2 \\
= \;& \sup_t \Big\| \, w(\sum_{j=1}^{t} \xi_j \phi(z_{-j}))^2 [\sum_{j=1}^{r} c_j I(u_{0j}(\theta_0) \le v_j) - G(v|\mathcal{Z}_{-t}^{-1},\theta_0)]^2 \\
& -E\{w(\sum_{j=1}^{t} \xi_j \phi(z_{-j}))^2 [\sum_{j=1}^{r} c_j I(u_{0j}(\theta_0) \le v_j) - G(v|\mathcal{Z}_{-t}^{-1},\theta_0)]^2 |\mathcal{Z}_{-m}^{m}\} \, \Big\|_2 \\
\le \;& w_{\max}^2 \sup_t \Big\| \, [\sum_{j=1}^{r} c_j I(u_{0j}(\theta_0) \le v_j) - G(v|\mathcal{Z}_{-t}^{-1},\theta_0)]^2 - E\{[\sum_{j=1}^{r} c_j I(u_{0j}(\theta_0) \le v_j) - G(v|\mathcal{Z}_{-t}^{-1},\theta_0)]^2 |\mathcal{Z}_{-m}^{m}\} \, \Big\|_2 \qquad \text{(A.31)} \\
& +2\tilde{w} \sum_{j=m+1}^{\infty} |\xi_j| \, \Big\| \, |\phi(z_{-j})| [\sum_{j=1}^{r} c_j I(u_{0j}(\theta_0) \le v_j) - G(v|\mathcal{Z}_{-t}^{-1},\theta_0)] \\
& -E\phi(z_{-j}) [\sum_{j=1}^{r} c_j I(u_{0j}(\theta_0) \le v_j) - G(v|\mathcal{Z}_{-t}^{-1},\theta_0)] \, \Big\|_2 , \qquad \text{(A.32)}
\end{aligned}$$

The first equality follows from the strict stationarity, and the last inequality from the mean value theorem. (A.31) is of size $-1/2$ by Assumption 1(b). (A.32) is of size $-1/2$ by $\sup_{\xi\in\Xi} \sum_{j=m+1}^{\infty} |\xi_j| = O(m^{-\zeta})$ by Assumption 1(e). Thus, $\{q_t(\theta_0,\gamma)^2\}_{t=1}^{\infty}$ is L^2-NED on $\{z_t\}$ of size $-1/2$. Therefore, Lemma A.5 is proved.

Q.E.D.

Proof of Theorem 3.1(b): The proof of Theorem 3.1(b) consists of the following three lemmas.

LEMMA A.6:

$$p \lim_{T\to\infty} \sup_{\gamma\in\Gamma} \left| \frac{1}{T} \sum_{t=1}^{T} E q_t(\theta_1,\gamma) \right| > 0, \tag{A.33}$$

LEMMA A.7:

$$\frac{1}{T} \sum_{t=1}^{T} [q_t(\theta_1,\gamma) - E q_t(\theta_1,\gamma)] = o_p(1), \tag{A.34}$$

LEMMA A.8:

$$\frac{1}{T} \sum_{t=1}^{T} [q_t(\hat{\theta}_T,\gamma) - q_t(\theta_1,\gamma)] = o_p(1). \tag{A.35}$$

Before we prove these lemmas, we shall briefly sketch the proof of Theorem 3.1(b).

$$T^{-\frac{1}{2}} Q_T(\hat{\theta}_T,\gamma) = \frac{1}{T} \sum_{t=1}^{T} q_t(\hat{\theta}_T,\gamma) = \frac{1}{T} \sum_{t=1}^{T} q_t(\theta_1,\gamma) + o_p(1) = \frac{1}{T} \sum_{t=1}^{T} E q_t(\theta_1,\gamma) + o_p(1) = O_p(1) \ (\ne o_p(1)).$$

The second equality follows from Lemma A.8, the third from Lemma A.7, and the fourth from Lemma A.6 and the continuity of $Eq_t(\theta_1,\gamma)$.

Proof of Lemma A.6:

$$\Big| \, \frac{1}{T} \sum_{t=1}^{T} E q_t(\theta_1,\gamma) - E w(\sum_{j=1}^{\infty} \xi_j \phi(z_{t-j})) [H(v|\mathcal{Z}_{-\infty}^{t-1}) - G(v|\mathcal{Z}_{-\infty}^{t-1},\theta_1)] \, \Big|$$

$$
\begin{aligned}
= & \ \Big| \frac{1}{T}\sum_{t=1}^{T} w(\sum_{j=1}^{t}\xi_j\phi(z_{-j}))[H(v|\mathcal{Z}_{-t}^{-1}) - G(v|\mathcal{Z}_{-t}^{-1},\theta_1)] - Ew(\sum_{j=1}^{\infty}\xi_j\phi(z_{-j}))[H(v|\mathcal{Z}_{-\infty}^{-1}) - G(v|\mathcal{Z}_{-\infty}^{-1},\theta_1)] \Big| \\
\leq & \ \Big| \frac{1}{T}\sum_{t=1}^{T} E[w(\sum_{j=1}^{t}\xi_j\phi(z_{-j})) - w(\sum_{j=1}^{\infty}\xi_j\phi(z_{-j}))][H(v|\mathcal{Z}_{-t}^{-1}) - G(v|\mathcal{Z}_{-t}^{-1},\theta_1)] \Big| \\
& + \Big| \frac{1}{T}\sum_{t=1}^{T} E\{w(\sum_{j=1}^{\infty}\xi_j\phi(z_{-j}))\left[H(v|\mathcal{Z}_{-t}^{-1}) - G(v|\mathcal{Z}_{-t}^{-1},\theta_1) - H(v|\mathcal{Z}_{-\infty}^{-1}) + G(v|\mathcal{Z}_{-\infty}^{-1},\theta_1)\right] \Big| \\
\leq & \ \frac{2|c|\tilde{w}}{T}\sum_{t=1}^{T}\sum_{j=t+1}^{\infty}|\xi_j|^2 j^{1+\zeta}\sum_{j=t+1}^{\infty}E|\phi(z_{-j})|^2 j^{-1-\zeta} \\
& + \frac{w_{\max}}{T}\sum_{t=1}^{T}\left\| H(v|\mathcal{Z}_{-t}^{-1}) - G(v|\mathcal{Z}_{-t}^{-1},\theta_1) - H(v|\mathcal{Z}_{-\infty}^{-1}) + G(v|\mathcal{Z}_{-\infty}^{-1},\theta_1)\right\|_2 . \qquad \text{(A.36)}
\end{aligned}
$$

The first term goes to zero uniformly in $\gamma \in$, since

$$
\lim_{T\to\infty}\sum_{t=1}^{T} t^{-1}\sum_{j=t+1}^{\infty}|\xi_j| j^{1+\zeta}\sum_{j=t+1}^{\infty}E|\phi(z_{-j})| j^{-1-\zeta} < \infty. \qquad \text{(A.37)}
$$

The second term goes to zero uniformly in $\gamma \in$, by Assumptions 1(b)(g). Thus,

$$
\lim_{T\to\infty}\sup_{\gamma\in\Gamma}\left|\frac{1}{T}\sum_{t=1}^{T}Eq_t(\theta_1,\gamma)\right| = \sup_{\gamma\in\Gamma}E\left\{w(\sum_{j=1}^{\infty}\xi_j\phi(z_{-j}))[H(v|\mathcal{Z}_{-\infty}^{-1}) - G(v|\mathcal{Z}_{-\infty}^{-1})]\right\} > 0. \qquad \text{(A.38)}
$$

The proof of the last inequality is analogous to that of Theorem 2 in De Jong (1996, pp.18-19) except that his Theorem 1 in the proof is replaced by Theorem 2.3 in Stinchcombe and White (1995). *Q.E.D.*

Proof of Lemma A.7: The left-hand side of (A.34) is less than or equal to

$$
\begin{aligned}
& \sup_{\gamma\in\Gamma}\Big| \frac{1}{T}\sum_{t=1}^{T} w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))[\sum_{j=1}^{r}c_j I(u_{tj}(\theta_1)\leq v_j) - G(v|\mathcal{Z}_0^{t-1},\theta_1)] \\
& -E\{w(\sum_{j=1}^{t}\xi_j\phi(z_{-j}))[H(v|\mathcal{Z}_{-t}^{-1}) - G(v|\mathcal{Z}_{-t}^{-1},\theta_1)]\} \Big| \\
\leq & \sup_{\gamma\in\Gamma}\Big| \frac{1}{T}\sum_{t=1}^{T} w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))[\sum_{j=1}^{r}c_j I(u_{tj}(\theta_1)\leq v_j) - H(v|\mathcal{Z}_0^{t-1})] \Big| \qquad \text{(A.39)} \\
& +\sup_{\gamma\in\Gamma}\Big| \frac{1}{T}\sum_{t=1}^{T} w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))[H(v|\mathcal{Z}_0^{t-1}) - G(v|\mathcal{Z}_0^{t-1},\theta_1)] \\
& -E\{w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))[H(v|\mathcal{Z}_0^{t-1}) - G(v|\mathcal{Z}_0^{t-1},\theta_1)]\} \Big| . \qquad \text{(A.40)}
\end{aligned}
$$

By the uniform SLLN, it suffices to show the total boundedness of (, , ρ), the stochastic equicontinuity of

$$
\{w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))[\sum_{j=1}^{r}c_j I(u_{tj}(\theta_1)\leq v_j) - H(v|\mathcal{Z}_0^{t-1})] : \gamma\in\ ,\ \}_{t=1}^{\infty}
$$

and

$$
\{w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))[H(v|\mathcal{Z}_0^{t-1}) - G(v|\mathcal{Z}_0^{t-1},\theta_1)]\}_{t=1}^{\infty},
$$

and the pointwise convergence. By Lemma A.1, $(, , \rho)$ is totally bounded. The proof of the stochastic equicontinuity is analogous to that of Lemma A.3 and thus is omitted. By Theorem 2.15 in Hall and Heyde (1980),

$$\Big| \sum_{t=1}^{T} t^{-1} w(\sum_{j=1}^{\infty} \xi_j \phi(z_{t-j}))[\sum_{j=1}^{r} c_j I(u_{tj}(\theta_1) \leq v_j) - H(v|\mathcal{Z}_0^{t-1})] \Big| < \infty \text{ a.s.} \tag{A.41}$$

By the Kronecker lemma,

$$\Big| \frac{1}{T} \sum_{i=1}^{T} w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))[\sum_{j=1}^{r} c_j I(u_{tj}(\theta_1) \leq v_j) - H(v|\mathcal{Z}_0^{t-1})] \Big| = o_{as}(1). \tag{A.42}$$

By Assumptions 1(a)(b)(g) and Theorem 3.1 in McLeish (1975),

$$\begin{aligned} &\frac{1}{T} \sum_{t=1}^{T} w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))[H(v|\mathcal{Z}_0^{t-1}) - G(v|\mathcal{Z}_0^{t-1}, \theta_1)] \\ &-E\{w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))[H(v|\mathcal{Z}_0^{t-1}) - G(v|\mathcal{Z}_0^{t-1}, \theta_1)]\} \quad = \quad o_{as}(1). \end{aligned} \tag{A.43}$$

Thus, both (A.39) and (A.40) are $o_{as}(1)$. The proof of Lemma A.7 is completed. *Q.E.D.*

Proof of Lemma A.8: The proof is similar to that of Lemma A.3 except that Assumption 1(d') is used instead of (3.8) and (3.9). Therefore, it is omitted. *Q.E.D.*

Proof of Theorem 3.2(a): By Lemma A.2, it suffices to show that

$$\begin{aligned} &\frac{1}{T} \sum_{t=1}^{T} \{[w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))(\sum_{j=1}^{r} c_j I(u_{tj}(\hat{\theta}_T) \leq v_j) - G(v|\mathcal{Z}_0^{t-1}, \hat{\theta}_T)) + D(\gamma)\psi(z_t, ..., z_0, \hat{\theta}_T)]^2 \\ &-[w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))(\sum_{j=1}^{r} c_j I(u_{tj}(\theta_0) \leq v_j) - G(v|\mathcal{Z}_0^{t-1}, \theta_0)) + D(\gamma)\psi_t(z_t, ..., z_0, \theta_0)]^2\} = o_p(1), \end{aligned} \tag{A.44}$$

and

$$\frac{1}{T} \sum_{t=1}^{T} \left[w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))(\sum_{j=1}^{r} c_j I(u_{tj}(\theta_0) \leq v_j) - G(v|\mathcal{Z}_0^{t-1}, \theta_0)) + D(\gamma)\psi_t(z_t, ..., z_0, \theta_0) \right]^2 - V(\gamma) = o_p(1), \tag{A.45}$$

where $D(\gamma) = E\{w(\sum_{j=1}^{\infty} \xi_j \phi(z_{-j}))[G_v(v|\mathcal{Z}_{-\infty}^{-1}, \theta_0) g_\theta(z_0, \theta_0) + G_\theta(v|\mathcal{Z}_{-\infty}^{-1}, \theta_0)]\}$. By the Cauchy-Schwarz inequality,

$$\begin{aligned} &\sup_{\gamma \in \Gamma} \Big| \frac{1}{T} \sum_{t=1}^{T} \{[w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))(\sum_{j=1}^{r} c_j I(u_{tj}(\hat{\theta}_T) \leq v_j) - G(v|\mathcal{Z}_0^{t-1}, \hat{\theta}_T)) + D(\gamma)\psi(z_t, ..., z_0, \hat{\theta}_T)]^2 \\ &-[w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))(\sum_{j=1}^{r} c_j I(u_{tj}(\theta_0) \leq v_j) - G(v|\mathcal{Z}_0^{t-1}, \theta_0)) + D(\gamma)\psi_t(z_t, ..., z_0, \theta_0)]^2\} \Big| \\ \leq\ &\sup_{\gamma \in \Gamma} \Big\{ \frac{1}{T} \sum_{t=1}^{T} [w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))(\sum_{j=1}^{r} c_j I(u_{tj}(\hat{\theta}_T) \leq v_j) - G(v|\mathcal{Z}_0^{t-1}, \hat{\theta}_T)) + D(\gamma)\psi_t(z_t, ..., z_0, \hat{\theta}_T) \\ &+w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))(\sum_{j=1}^{r} c_j I(u_{tj}(\theta_0) \leq v_j) - G(v|\mathcal{Z}_0^{t-1}, \theta_0)) + D(\gamma)\psi_t(z_t, ..., z_0, \theta_0)]^2 \Big\}^{\frac{1}{2}} \end{aligned} \tag{A.46}$$

$$\times \sup_{\gamma \in \Gamma} \Big\{ \frac{1}{T} \sum_{t=1}^{T} [w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))(\sum_{j=1}^{r} c_j I(u_{tj}(\hat{\theta}_T) \leq v_j) - G(v|\mathcal{Z}_0^{t-1}, \hat{\theta}_T)) + D(\gamma)\psi_t(z_t, ..., z_0, \hat{\theta}_T)$$

$$-w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))(\sum_{j=1}^{r} c_j I(u_{tj}(\theta_0) \leq v_j) - G(v|\mathcal{Z}_0^{t-1}, \theta_0)) - D(\gamma)\psi_t(z_t, ..., z_0, \theta_0)]^2 \Big\}^{\frac{1}{2}}. \qquad \text{(A.47)}$$

By Assumptions 1(a)(b)(d) and Theorem 3.1 in McLeish (1975), one can show that (A.46) is $O_{as}(1)$. Thus, it suffices to show that (A.47) is $o_{as}(1)$.

$$\frac{1}{T} \sum_{t=1}^{T} [w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))(\sum_{j=1}^{r} c_j I(u_{tj}(\hat{\theta}_T) \leq v_j) - G(v|\mathcal{Z}_0^{t-1}, \hat{\theta}_T)) + D(\xi)\psi(z_t, ..., z_0, \hat{\theta}_T)$$

$$-w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))(\sum_{j=1}^{r} c_j I(u_{tj}(\theta_0) \leq v_j) - G(v|\mathcal{Z}_0^{t-1}, \theta_0)) - D(\xi)\psi_t(z_t, ..., z_0, \theta_0)]^2$$

$$\leq \quad \frac{4w_{\max}}{T} \sum_{t=1}^{T} \sum_{j=1}^{r} c_j^2 [I(u_{tj}(\hat{\theta}_T) \leq v_j) - I(u_{tj}(\theta_0) \leq v_j)]^2 \qquad \text{(A.48)}$$

$$+\frac{2w_{\max}}{T} \sum_{t=1}^{T} \bar{G}_t(z_{t-1}, z_{t-2}, ..., z_0)|\hat{\theta}_T - \theta_0| \qquad \text{(A.49)}$$

$$+\frac{4w_{\max}}{T} \sum_{t=1}^{T} D(\xi)^2 \bar{\psi}_t(z_t, ..., z_0)|\hat{\theta}_T - \theta_0|^2. \qquad \text{(A.50)}$$

A similar technique used in the proof of Lemma A.3 proves that (A.48) is $o_{as}(1)$. (A.49) and (A.50) are $o_{as}(1)$ uniformly in $v \in \Upsilon$ by Assumptions 1(a)(b)(d) and Theorem 3.1 in McLeish (1975).

The second part of Theorem 3.2(a) is straightforward and thus is omitted. *Q.E.D.*

Proof of Theorem 3.2(b): To simplify notation, let $r = 1$. It suffices to show

$$\frac{1}{2h_T T} \sum_{t=1}^{T} w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))\hat{G}_{vt}(\hat{\theta}_T) g_\theta(z_t, \hat{\theta}_T) - \frac{1}{2h_T T} \sum_{t=1}^{T} w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))\hat{G}_{vt}(\theta_0) g_\theta(z_t, \theta_0) = o_p(1), \qquad \text{(A.51)}$$

$$\frac{1}{2h_T T} \sum_{t=1}^{T} w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))\hat{G}_{vt}(\theta_0) g_\theta(z_t, \theta_0)$$

$$-\frac{1}{2h_T T} \sum_{t=1}^{T} w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))[G(v + h_T|\mathcal{Z}_0^{t-1}, \theta_0) - G(v - h_T|\mathcal{Z}_0^{t-1}, \theta_0)] g_\theta(z_t, \theta_0) = o_p(1), \qquad \text{(A.52)}$$

and

$$\frac{1}{2h_T T} \sum_{t=1}^{T} w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j}))[G(v + h_T|\mathcal{Z}_0^{t-1}) - G(v - h_T|\mathcal{Z}_0^{t-1})] g_\theta(z_t, \theta_0)$$

$$-\frac{1}{T} \sum_{t=1}^{T} E[w(\sum_{j=1}^{t} \xi_j \phi(z_{t-j})) G_v(v|\mathcal{Z}_0^{t-1}, \theta_0) g_\theta(z_t, \theta_0)] = o_p(1). \qquad \text{(A.53)}$$

Since the proofs of (A.51), (A.52), and (A.53) are similar to those of Lemmas A.2, A.3, and A.4, we only sketch them.

As in the proof of Lemma A.3, one can show that

$$\frac{1}{h_T T}\sum_{t=1}^{T} w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))\hat{G}_{vt}(\hat{\theta}_T)g_\theta(z_t,\hat{\theta}_T) - \frac{1}{Th_T}\sum_{t=1}^{T} w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))\hat{G}_{vt}(\theta_0)g_\theta(z_t,\theta_0) = O_p\left(\frac{\hat{\theta}_T-\theta_0}{h_T}\right). \tag{A.54}$$

Since $h_T T^{1/2} \to \infty$ as $T \to \infty$, (A.51) follows.

The proof of (A.52) consists of the proofs of the total boundedness, stochastic equicontinuity, and pointwise convergence. By Lemma A.1, $(,\ ,\rho)$ is totally bounded. The proof of stochastic equicontinuity is analogous to that of Lemma A.4 and thus is omitted. Since

$$\Big\{ w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j}))\hat{G}_{vt}(\theta_0)g_\theta(z_t,\theta_0) - \frac{1}{2h_T T}\sum_{t=1}^{T} w(\sum_{j=1}^{t}\xi_j\phi(z_{t-j})) \times[G(v+h_T|\mathcal{Z}_0^{t-1},\theta_0) - G(v-h_T|\mathcal{Z}_0^{t-1},\theta_0)]g_\theta(z_t,\theta_0)\Big\}_{T=1}^{\infty} \tag{A.55}$$

is a martingale difference array, one can show the pointwise convergence by Theorem 2 in Andrews (1988a).

Since

$$\lim_{h_T\to 0}\frac{G(v+h_T|z_{t-1},...,z_0) - G(v-h_T|z_{t-1},...,z_0)}{2h_T} = G_v(v|z_{t-1},...,z_0) \text{ a.s.} \tag{A.56}$$

the proof of (A.53) is straightforward. *Q.E.D.*

References

Andrews, D.W.K. (1988a) "Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables," *Econometric Theory,* 4, 458–467.

——— (1988b) "Chi-Square Diagnostic Tests for Econometric Models: Theory," *Econometrica,* 56, 1419–1453.

——— (1997) "A Conditional Kolmogorov Test," *Econometrica,* 64, 1097–1128.

Bai, J. (1996) "Testing for Parameter Constancy in Linear Regressions: An Empirical Distribution Function Approach," *Econometrica,* 64, 597–622.

——— (1997) "Testing Parametric Conditional Distributions of Dynamic Models," manuscript, Massachusetts Institute of Technology.

Bickel, P.J. (1982) "On Adaptive Estimation," *Annals of Statistics,* 10, 647–671.

Bierens, H.J. (1990) "A Consistent Conditional Moment Test of Functional Form," *Econometrica,* 58, 1443–1458.

Bierens, H.J. and D. Ginther (1997) "Integrated Conditional Moment Estimator and Testing of Median Regression Models with an Application to a Mincer-type Wage Equation," Pennsylvania State University and Southern Methodist University, manuscript.

Bierens, H.J. and W. Ploberger (1997) "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica,* 63, 1129–1151.

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

Christoffersen, P.F. (1997) "Evaluating Interval Forecasts," *International Economic Review,* forthcoming.

De Jong, R.M. (1996) "The Bierens Test Under Data Dependence," *Journal of Econometrics,* 72, 1–32.

Diebold, F.X., T. Gunther, and A.S. Tay (1997) "Evaluating Density Forecasts," *International Economic Review,* forthcoming.

Fan, Y. and R. Gencay (1995) "A Consistent Nonparametric Test of Symmetry in Linear Regression Models," *Journal of the American Statistical Association,* 90, 551–557.

Gallant, A.R. and H. White (1988) *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models,* New Basil Blackwell: New York.

Hall, P. and C.C. Heyde (1980) *Martingale Limit Theory and its Application,* Academic Press: New York.

Heckman, J.J. (1984) "The χ^2 Goodness of Fit Statistic for Models with Parameters Estimated From Microdata," *Econometrica,* 52, 1543–1547.

Koenker, R. and G. Bassett, Jr. (1978) "Regression Quantiles," *Econometrica,* 46, 33–50.

Lehmann, E.C. (1986) *Testing Statistical Hypotheses,* 2nd ed., John WIley & Sons: New York.

Linton, O. and P. Gozalo (1996) "Conditional Independence Restrictions: Testing and Estimation," manuscript, Yale University and Brown University.

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.

McLeish, D.L. (1975) "A Maximal Inequality and Dependent Strong Laws," *Annals of Probability,* 3, 826–836.

Newey, W.K. (1988) "Adaptive Estimation of Regression Models via Moment Restrictions," *Journal of Econometrics,* 38, 301–339.

Newey, W.K. and D.G. Steigerwald (1997) "Asymptotic Bias for Quasi-Maximum-Likelihood Estimators in Conditional Heteroscedasticity Models," *Econometrica,* 65, 587–599.

Stinchcombe, M.B. and H. White (1998) "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative," *Econometric Theory,* 14, 295–325.

Zheng, J.X. (1994) "A Consistent Test of Conditional Parametric Distributions," manuscript, University of Texas, Austin.

——— (1996) "A Specification Test of Conditional Parametric Distributions using Kernel Estimation Methods," manuscript, University of Texas, Austin.

——— (1998a) "A Consistent Nonparametric Test of Parametric Regression Models Under Conditional Quantile Restrictions," *Econometric Theory,* 14, 123–138.

——— (1998b) "Consistent Specification Testing for Conditional Symmetry," *Econometric Theory,* 14, 139–149.

Table 1

Finite-Sample Performance of Tests for Conditional Goodness-of-Fit Restrictions

$\mathrm{H}_0: \ y_t = \varepsilon_t, \ \varepsilon_t \sim NID(\mu, \sigma^2)$ for some μ and σ^2

DGP1			$y_t = \varepsilon_t, \varepsilon_t \sim NID(0,1)$					
T=250			T=500			T=1000		
1%	5%	10%	1%	5%	10%	1%	5%	10%
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
DGP2			$y_t = \varepsilon_t^2 - 1, \varepsilon_t \sim NID(0,1)$					
T=250			T=500			T=1000		
1%	5%	10%	1%	5%	10%	1%	5%	10%
99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
DGP3	$y_t = \exp(-h_t/2)\varepsilon_t, h_t = 0.10 + 0.90 h_{t-1} + \sqrt{0.3}\eta_t (\varepsilon_t, \eta_t) \sim NID(0, I_2)$							
T=250			T=500			T=1000		
1%	5%	10%	1%	5%	10%	1%	5%	10%
0.1	2.8	7.0	4.4	26.0	48.9	49.5	91.3	97.9
DGP4			$y_t = 0.5 y_{t-1} + \varepsilon_t, \ \varepsilon_t \sim NID(0,1)$					
T=250			T=500			T=1000		
1%	5%	10%	1%	5%	10%	1%	5%	10%
0.0	0.1	1.0	0.0	12.2	63.7	39.1	99.6	100.0
DGP5			$y_t = 0.9 y_{t-1} + \varepsilon_t, \ \varepsilon_t \sim NID(0,1)$					
T=250			T=500			T=1000		
1%	5%	10%	1%	5%	10%	1%	5%	10%
4.1	93.1	99.3	100.0	100.0	100.0	100.0	100.0	100.0

Notes: The Monte Carlo integration method is used to evaluate the test statistics with the number of repetitions set to 300. The number of Monte Carlo replications is to 1000.