A Mean Likelihood Ratio Specification Test

Isaiah Andrews, MIT*

December 21, 2012

Abstract

This paper considers the problem of specification testing in general parametric models and shows that for a wide class of models the hypothesis of correct specification is equivalent to a continuum of moment equalities. Using these moment equalities we construct a class of specification tests that have correct asymptotic size in general parametric models, including stationary time series models, and that are consistent when the above equivalence holds. We show that the proposed tests have power against $\sqrt{T}$-local alternatives and compare them to previously proposed consistent tests of distributional specification, both from a theoretical perspective and in simulation.

JEL Classification: C52

Keywords: Consistent Specification Testing, Time Series

Specification testing has long been a central topic of research in econometrics. Early work by Hausman (1978) compares two different estimators which have the same limit under the null of correct specification and rejects if they differ by too large a margin. White (1980) and (1982) both compare different variance matrix estimators, the former to detect heteroskedasticity and the latter to detect misspecification in a general parametric setting. Newey (1985) considers tests of conditional moment restrictions and shows that both the Hausman and White tests can be considered as special cases of this larger class. More recently, Chesher et. al. (1999) consider what they call Bartlett Identities Tests, analogs of the White (1982) Information

*Department of Economics, Massachusetts Institute of Technology, 50 Memorial Drive Cambridge MA 02142 USA. Phone: 617-835-5274, Email: iandrews@mit.edu

Matrix test for parametric models which test higher order restrictions imposed by the null of correct specification.

All of these tests consider finite collections of restrictions implied by the hypothesis of correct specification. None of these tests is consistent against all possible alternatives in general models, since it is possible to find a data generating process which differs from the null but still satisfies the tested restrictions. Motivated by this fact, we propose a test which is consistent against fixed alternatives in models with independent observations and against a broad class of alternatives in models with dependent observations, and which has power against $\sqrt{T}$ local alternatives. This test complements the existing econometric literature on consistent tests of distributional specification, including (D.) Andrews (1997), Zheng (2000), and Bierens and Wang (2012).

Andrews (1997) proposes an extension of the classical Kolmogorov-Smirnov statistic for testing conditional distributional specification in independent data. The proposed test is shown to be consistent against all fixed alternatives and to have nontrivial power against $\sqrt{T}$ local alternatives. Zheng (2000) offers an alternative approach to specification testing, based on an approximation to the average Kullback-Leibler divergence between the data and the null parametric model. This approach requires that the modeled variable X be continuously distributed and, due to its use of a kernel density estimator, has power against local alternatives only if they approach the null at a rate slower than $\sqrt{T}$. A recent paper by Bierens and Wang (2012) takes a different approach to the problem, comparing the model implied characteristic function with the characteristic function of the empirical distribution, which they show yields a test that is consistent against fixed alternatives and has power against $\sqrt{T}$ local alternatives.

A difference between the approach considered here and those of Andrews (1997), Zheng (2000), and Bierens and Wang (2012) is that the statistic we consider can also be used to test the null of correct specification in contexts with dependent observations, such as time series. The tests proposed by Hausman (1978), White (1982), and Newey (1985) can all be applied to dependent data, as can more recent procedures by Inoue (1997), who considers testing linear restrictions on the conditional distribution, and Bai (2003) who considers a test for distributional specification in general time series models. None of these procedures is consistent, however, while our test is consistent provided both the null and alternative distributions depend on the past only through a finite-dimensional vector of state variables. The problem of consistent testing of distributional specification with dependent data is considerably more difficult than in the independent case, however, and as discussed below the approach proposed here is not in general consistent against unrestricted dependent alternatives.

For the remainder of the paper, we assume that we observe a sample $\{(X_t, Z_t)\}_{t=1}^T$ and denote by $\mathcal{F}_t$ the $\sigma-$algebra generated by $\{(X_s, Z_{s+1})\}_{s=1}^t$ (so Z_t is measurable with respect to $\mathcal{F}_{t-1}$). Suppose we think that conditional on $\mathcal{F}_{t-1}$, X_t is distributed according to $F(x|\theta_0, \mathcal{F}_{t-1})$ for some (in general unknown) parameter θ_0 in some parameter space $\Theta \subset \mathbb{R}^k$. We propose a class of tests for the hypothesis of correct parametric specification

$$X_t|\mathcal{F}_{t-1} \sim F(x|\theta_0, \mathcal{F}_{t-1}) \, a.s. \text{ for some } \theta_0 \in \Theta$$

and show that these tests control size (where a.s. stands for almost surely). We are particularly interested in null hypotheses of the form $H_0 : X_t|\mathcal{F}_{t-1} \sim F(x|\theta_0, Z_t) \, a.s.$, where the distribution of X_t depends on the past only through a known, finite-dimensional vector of conditioning variables, as these are the null hypotheses for which we can obtain consistency results. For this reason, and to simplify notation, we focus on the case $F(x|\theta_0, \mathcal{F}_{t-1}) = F(x|\theta_0, Z_t)$ for the remainder of the analysis, though our results concerning asymptotic size hold even for $F(x|\theta_0, \mathcal{F}_{t-1}) \neq F(x|\theta_0, Z_t)$.

We propose tests based on Mean Likelihood Ratio (MLR) statistics of the form

$$R(\hat{\theta}, \lambda, \xi) = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{g(X_t|\lambda, Z_t)}{f(X_t|\hat{\theta}, Z_t)} - 1 \right) \psi(\Phi(Z_t)'\xi)$$

where $\hat{\theta}$ is an estimator of θ_0, g is a function (typically a density or probability mass function) which integrates to one, f is the density or probability mass function of F, Φ is a bounded one-to-one function, and ψ is a non-polynomial monotone analytic function. Under the null of correct specification and some further assumptions we show that $R(\hat{\theta}, \lambda, \xi) \to_p 0$ uniformly in $(\lambda, \xi) \in \Lambda \times \Xi$ for Λ, Ξ compact, and so base our test on $\sup_{(\lambda,\xi)\in\Lambda\times\Xi} |R(\hat{\theta}, \lambda, \xi)|$. We show that statistics of this form have several interesting properties: first as mentioned above, they may be used to construct tests in both independent and dependent data models. Moreover, the resulting tests are consistent against all fixed (global) alternatives in independent models and against alternatives which imply a different stationary distribution of the modeled variables

X_t and the (finite) vector of conditioning variables Z_t in dependent models. To clarify this last point, consider the modified null hypothesis

$$H_0^* : X_t|Z_t \sim F(x|\theta_0, Z_t)\, a.s.$$

which restricts only the stationary distribution of (X_t, Z_t), rather than the distribution of $X_t|\mathcal{F}_{t-1}$. We show that appropriately constructed MLR tests are consistent against alternatives which violate H_0^*. In independent models (that is models where (X_t, Z_t) is independent of (X_s, Z_s) $\forall\ t \neq s$), H_0^* and H_0 are equivalent so this implies the consistency of MLR tests against all fixed alternatives in models with independent data. In dependent models, however, $H_0 \subset H_0^*$ so while we prove consistency against alternatives which violate H_0^*, there are models in $H_0^* \backslash H_0$ which violate the tested null (H_0) against which the MLR test is not consistent.

We show that like the tests proposed by Andrews (1997) and by Bierens and Wang (2012), MLR tests have power against $\sqrt{T}$ local alternatives. The proposed tests perform well in simulation, controlling size in both independent and dependent models while offering power competitive with that of existing procedures in independent models. Different tests in the family we consider have power against different alternatives in finite samples so, as we illustrate in simulation, if one is especially concerned about particular forms of misspecification one can choose tests with more power in the appropriate directions.

We begin in Section 1 by showing that the restriction $H_0^* : X_t|Z_t \sim F(x|\theta_0, Z_t)\, a.s.$ is equivalent to a continuum of conditional moment equalities, which are in turn equivalent to a continuum of unconditional moment equalities that we use to construct our family of specification tests. In Section 2 we propose a test statistic for the hypothesis $H_0 : X_t|\mathcal{F}_{t-1} \sim F(x|\theta_0, Z_t)\, a.s.$, derive its limiting distribution and prove that tests based on bootstrap critical values have correct asymptotic size. In Section 3 we prove that for appropriate choices of $g(x|\lambda, z)$, MLR tests of H_0 are consistent against alternatives which violate H_0^* and have nontrivial power against $\sqrt{T}$ local alternatives. Section 4 compares the proposed tests to others in the literature from a theoretical perspective, while simulation results are presented in Section 5. All proofs may be found in the Appendix.

1 Equivalence of Restrictions

We begin by observing that $H_0^* : X_t|Z_t \sim F(x|\theta_0, Z_t)\, a.s.$ is equivalent to a continuum of moment conditions. For convenience (and without loss of generality), we assume that Z_t includes a constant term. Suppose that for all $\theta \in \Theta$ and all values of Z_t, $F(x|\theta, Z_t)$ is absolutely continuous with respect to some σ-finite measure μ, and denote its Radon-Nikodym derivative with respect to μ by $f(x|\theta, Z_t)$.

Let $\{G(x|\lambda, z) : \lambda \in \Lambda, z \in \text{Supp}(Z_t)\}$, for $\text{Supp}(Z_t)$ the support of Z_t, be a family of (possibly signed) measures parametrized by $\lambda \in \Lambda$. We take Λ to be a compact set with a nonempty interior and assume that $G(x|\lambda, z)$ and $F(x|\theta, z)$ are mutually absolutely continuous for all $\lambda \in \Lambda$, $\theta \in \Theta$, and $z \in \text{Supp}(Z_t)$. We denote the Radon-Nikodym derivative of $G(x|\lambda, z)$ with respect to μ by $g(x|\lambda, z)$ and assume that $\int g(x|\lambda, z) d\mu = 1$.

We first show that for fixed z and appropriately chosen g, the restriction $X_t|Z_t = z \sim F(x|\theta_0, z)$ is equivalent to a continuum of conditional moment equalities:

$$E_{f_0}\left[\left.\frac{g(X_t|\lambda, z)}{f(X_t|\theta_0, z)}\right| Z_t = z\right] = 1 \;\forall \lambda \in \Lambda. \tag{1}$$

Note that

$$E_{f(\theta_0)}\left[\left.\frac{g(X_t|\lambda, z)}{f(X_t|\theta_0, z)}\right| Z_t = z\right] = \int \frac{g(x|\lambda, z)}{f(x|\theta_0, z)} f(x|\theta_0, z) d\mu(x) = E_{g(\lambda)}\left[\left.\frac{f(X_t|\theta_0, z)}{f(X_t|\theta_0, z)}\right| Z_t = z\right] = 1,$$

so our distributional assumption implies a continuum of moment restrictions for each value of z (i.e. a restriction for each $\lambda \in \Lambda$). We next show that if the family $\{G(x|\lambda, z) : \lambda \in \Lambda\}$ is complete for all $z \in \text{Supp}(Z_t)$, this continuum of moment conditions is also sufficient for correct specification. A family $\mathcal{P}$ of measures is called complete if for any measurable function ϕ, $E_P[\phi(x)] = 0$ for all $P \in \mathcal{P}$ implies that $\phi(x) = 0$ almost surely with respect to all $P \in \mathcal{P}$ (see Lehman and Romano (2005), Section 4.3). The appropriate sufficient conditions for completeness will depend on the context under consideration. A useful result in Lehman and Romano (2005) (Theorem 4.3.1) is that it suffices for $g(x|\lambda, z)$ to be an exponential family distribution such that $g(x|\lambda, z) = h(x, z)a(\lambda, z)\exp(m(\lambda, z)'T(x, z))$ with $T(x, z)$ a one-to-one function of x and m a function such that the image of Λ under $m(\cdot, z)$ contains an open set.

Suppose that conditional on $Z_t = z$ the true distribution of X_t is $F_0(x|z)$ which is mutually absolutely continuous with respect to $G(\lambda, z)$ for all $\lambda \in \Lambda$, with Radon-Nikodym derivative $f_0(x|z)$ with respect to μ such that the continuum of moment conditions (1) holds. Hence,

$$E_{f_0}\left[\frac{g(X_t|\lambda,z)}{f(X_t|\theta_0,z)}\bigg| Z_t = z\right] - 1 = \int \frac{g(x|\lambda,z)}{f(x|\theta_0,z)} f_0(x|z)d\mu - 1$$

$$= \int \left(\frac{f_0(x|z)}{f(x|\theta_0,z)} - 1\right) g(x|\lambda,z)d\mu = 0.$$

If $g(x|\lambda, z)$ is a complete family, however, this condition can be satisfied for all λ if and only if $\frac{f_0(X_t|z)}{f(X_t|\theta_0,z)} = 1$ almost surely with respect to the true distribution $F_0(x|z)$. To state this formally, we begin by requiring the stationarity of the process $\{(X_t, Z_t)\}$ so that we can meaningfully discuss the stationary distributions F_{XZ}, F_X, and F_Z:

Assumption 1 *The process $\{(X_t, Z_t)_{t=1}^{\infty}\}$ is stationary, with stationary distribution F_{XZ} and marginal distributions F_X and F_Z.*

Assumption 2 *$\{G(x|\lambda, z) : \lambda \in \Lambda\}$ is a complete family for all fixed z and $E_{f_0}\left[\frac{g(X_t|\lambda,z)}{f(X_t|\theta_0,z)}\Big| Z_t = z\right]$ exists for all $\lambda \in \Lambda$ and $z \in$Supp(F_Z).*

Theorem 1 *Under Assumptions 1 and 2, for any fixed $z \in$Supp(Z_t) $E_{f_0}\left[\frac{g(X_t|\lambda)}{f(X_t|\theta_0,z)} - 1\Big| Z_t = z\right] = 0$ for all $\lambda \in \Lambda$ if and only if $f_0(X_t|z) = f(X_t|\theta_0, z)$ almost surely (with respect to the true measure $F_0(x|z)$).*

We've established that our original parametric restriction is equivalent to a continuum of conditional moment restrictions. Since such restrictions are in general difficult to test directly, we next discuss an equivalent continuum of unconditional moment restrictions, derived as in the literature on integrated conditional moment tests originated by Bierens (1982). Specifically, we make the following assumption:

Assumption 3 *$E_{f_0}\left[\frac{g(X_t|\lambda,Z_t)}{f(X_t|\theta_0,Z_t)} - 1\Big| Z_t\right]$ is continuous in λ almost surely with respect to F_Z.*

Assumption 3 ensures that the conditional expectation $E_{f_0}\left[\frac{g(X_t|\lambda,Z_t)}{f(X_t|\theta_0,Z_t)} - 1\Big| Z_t\right]$ is well-behaved as a function of λ. Let $\xi \in \Xi \subset \mathbb{R}^l$ for some compact set Ξ with positive Lebesgue measure. Combining the results of Stinchcombe and White (1997) with Theorem 1, we obtain the following:

Theorem 2 *For $\Phi(z)$ a bounded one-to-one mapping, $\psi(\cdot)$ a non-polynomial real-valued monotone analytic function, and g as defined above, provided that $E_{f_0}\left[\frac{g(X_t|\lambda,Z_t)}{f(X_t|\theta_0,Z_t)}-1\right]$ exists $\forall\lambda\in\Lambda$ and Assumptions 1-3 hold,*

$$E_{f_0}\left[\left(\frac{g(X_t|\lambda,Z_t)}{f(X_t|\theta_0,Z_t)}-1\right)\psi(\Phi(Z_t)'\xi)\right]=0\ \forall(\lambda,\xi)\in\Lambda\times\Xi, \tag{2}$$

if and only if $f_0(X_t|Z_t)=f(X_t|\theta_0,Z_t)$ almost surely with respect to the true stationary distribution F_{XZ}.

As the results above establish, for $g(x|\lambda)$ a complete family the restriction that $f_0(X_t|Z_t)=f(X_t|\theta_0,Z_t)$ almost surely is equivalent to the restriction (2). Hence, to test model specification it suffices to test this continuum of unconditional moment equalities.

2 The Proposed Test

In this section, we derive size results for tests of the null hypothesis

$$H_0: X_t|\mathcal{F}_{t-1}\sim F(x|\theta_0,Z_t)\, a.s. \text{ for some } \theta_0\in\Theta.$$

To this end we consider the statistic

$$R(\theta,\lambda,\xi)=\frac{1}{T}\sum_{t=1}^{T}r_t(\theta,\lambda,\xi)=\frac{1}{T}\sum_{t=1}^{T}\left(\frac{g(X_t|\lambda,Z_t)}{f(X_t|\theta,Z_t)}-1\right)\psi(\Phi(Z_t)'\xi).$$

Below, we derive the limiting distribution of $\sqrt{T}R(\hat\theta,\lambda,\xi)$ under the null of correct specification, where $\hat\theta$ is a well-behaved estimator of θ (for example the maximum likelihood estimator in most models). To test the null hypothesis we consider the largest value taken by this statistic, $||\sqrt{T}R(\hat\theta,\lambda,\xi)||_\infty=\sup_{(\lambda,\xi)\in\Lambda\times\Xi}|\sqrt{T}R(\hat\theta,\lambda,\xi)|$, which should be small if the null hypothesis is correct.[1]

In our exposition we focus on testing null hypotheses of the form $H_0: X_t|\mathcal{F}_{t-1}\sim F(x|\theta,Z_t)$. As noted in the introduction, however, the results of this section on the size of MLR tests

[1]While we focus on the sup norm of $\sqrt{T}R(\hat\theta,\lambda,\xi)$, it is straightforward to adapt the results presented below to other functionals, for example the L^2-norm $\left(\int\left(\sqrt{T}R(\hat\theta,\lambda,\xi)\right)^2dH(\lambda,\xi)\right)^{\frac{1}{2}}$ for some measure H on $\Lambda\times\Xi$.

extend directly to tests of the broader class of null hypotheses $X_t|\mathcal{F}_{t-1} \sim F(x|\theta, \mathcal{F}_{t-1})$ where we allow the conditional distribution of X_t under the null to depend on all of $\mathcal{F}_{t-1}$ rather than just the finite-dimensional vector Z_t. In particular, by replacing $F(x|\theta, Z_t)$ by $F(x|\theta, \mathcal{F}_{t-1})$, $E[\cdot|Z_t]$ by $E[\cdot|\mathcal{F}_{t-1}]$ and so on, our proofs can easily be adapted to show that tests based on $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{g(X_t|\lambda, \mathcal{F}_{t-1})}{f(X_t|\theta, \mathcal{F}_{t-1})} - 1 \right) \psi(\Phi(Z_t)'\xi)$ will have correct asymptotic size, though they will not be consistent without further restrictions.

2.1 Limiting Distribution

We begin by showing that under the null of correct specification $\sqrt{T} R\left(\hat{\theta}, \lambda, \xi\right)$, viewed as a random function of (λ, ξ), converges weakly to a mean-zero Gaussian process. In our analysis we allow for two distinct types of conditioning variables Z_t, which we denote $Z_t^{(1)}$ and $Z_t^{(2)}$ respectively. $Z_t^{(1)}$ can be written in terms of past values of (X_s, Z_s) (which to say it is measurable with respect to $\mathcal{F}_{t-1}$) while $Z_t^{(2)}$ contains exogenous variables which evolve independently of X_t. In particular, we impose that conditional on $\left\{Z_1^{(2)}, ...Z_{t-2}^{(2)}, Z_{t-1}^{(2)}, Z_t^{(2)}\right\}$, $\{X_1, ...X_{t-2}, X_{t-1}, X_t\}$ and $\left\{Z_{t+1}^{(2)}, Z_{t+2}^{(2)}, Z_{t+3}^{(2)}...\right\}$ are independent. Using this assumption, we conduct our analysis conditional on $\left\{Z_t^{(2)}\right\} = \left\{Z_t^{(2)}\right\}_{t=1}^{\infty}$. Hence, for example, we will assume that for some function h, $\frac{1}{T} \sum_{t=1}^{T} h(X_t, Z_t) \to_p h^*$ where the convergence in probability is with respect to the distribution of the process $\{(X_t, Z_t)\}_{t=1}^{T} \Big| \left\{Z_t^{(2)}\right\}_{t=1}^{\infty}$. Likewise, many of our assumptions will concern terms of the form $E\left[h(X_t, Z_t)|Z_t, \left\{Z_t^{(2)}\right\}\right]$, though since under the null this is equal to $E[h(X_t, Z_t)|Z_t]$, (see the proof of Lemma A.3 for further discussion of this point) we will generally prefer the latter formulation for the sake of brevity. We condition on $\left\{Z_t^{(2)}\right\}$ for a number of reasons: first, it allows us to treat the independent and dependent cases in a unified fashion and to use the same bootstrap approach for both. Further, it greatly simplifies the exposition and allows us to cover a wide range of possibilities for the evolution of the unmodeled conditioning variables $Z_t^{(2)}$. The cost of this approach is that we must state our assumptions conditional on $\left\{Z_t^{(2)}\right\} = \left\{Z_t^{(2)}\right\}_{t=1}^{\infty}$, which may seem less intuitive. However, in many contexts it is straightforward to give sufficient conditions to ensure that our assumptions hold conditional on $\left\{Z_t^{(2)}\right\}$ with probability one, as is done in Andrews (1997) for the independent case.

We begin by formalizing the exogenity assumption on $Z_t^{(2)}$:

Assumption 4 *Under H_0 we have that $Z_t^{(2)}$ is exogenous: that is that conditional on $\left\{Z_s^{(2)}\right\}_{s=1}^{t}$, $\{X_s\}_{s=1}^{t}$ and $\left\{Z_s^{(2)}\right\}_{s=t+1}^{\infty}$ are independent.*

Next, we assume that the conditional density of X_t given Z_t is well behaved as a function of θ:

Assumption 5 *The conditional density $f(X_t|\theta, Z_t)$ satisfies:*

1. *$f(X_t|\theta, z)$ differentiable in θ on a neighborhood N_1 of θ_0 for all $(x, z) \in Supp\,(X_t, Z_t)$.*
2. *For all sequences of positive constants δ_T such that $\delta_T \to 0$, under the null we have that conditional on $\left\{Z_t^{(2)}\right\}$,*

$$\sup_{\lambda,\xi} \sup_{\theta:||\theta-\theta_0||<\delta_T} \left\| \frac{1}{T}\sum_{t=1}^{T} \frac{\partial}{\partial\theta} f(X_t|\theta, Z_t) \frac{g(X_t|\lambda, Z_t)}{f^2(X_t|\theta, Z_t)} \psi(\Phi(Z_t)'\xi) - \Delta_0(\xi, \lambda) \right\| \to_p 0.$$

   where $\Delta_0(\lambda, \xi) = E\left[\frac{\partial}{\partial\theta} f(X_t|\theta_0, Z_t) \frac{g(X_t|\lambda,Z_t)}{f^2(X_t|\theta_0,Z_t)} \psi(\Phi(Z_t)'\xi)\right]$.
3. *$\sup_{\lambda,\xi} ||\Delta_0(\lambda, \xi)|| < \infty$ and Δ_0 is continuous on $\Lambda \times \Xi$.*

Assumption 5(1), differentiability of $f(X_t|Z_t, \theta)$, is generally innocuous and is satisfied in a wide range of models. Assumption 5(2) is somewhat stronger but again is not usually problematic, provided that the family $\{G(x|\lambda, z) : \lambda \in \Lambda\}$ which we choose is well-behaved. If Assumption 5(2) is satisfied then 5(3) is, again, fairly weak. In particular, if $\Delta_0(\lambda, \xi)$ is continuous in (λ, ξ) by the compactness of $\Lambda \times \Xi$ Assumption 5(3) is equivalent to point-wise finiteness of $||\Delta_0(\lambda, \xi)||$.

Our next assumption formalizes the requirement that $\hat{\theta}$ be "well-behaved", in particular requiring that it have a linear asymptotic expansion:

Assumption 6 *Conditional on $\left\{Z_t^{(2)}\right\}$ the estimator $\hat{\theta}$ satisfies:*

1. *Under the null $X_t|\mathcal{F}_{t-1} \sim F(x|Z_t, \theta_0)$, $\sqrt{T}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{T}} \sum_t I_0 \gamma(X_t, Z_t, \theta_0) + o_p(1)$ where I_0 is a k-dimensional nonrandom matrix which may depend on θ_0.*
2. *$\gamma(x, z, \theta)$ is a measurable function satisfying*

   (a) $E\left[\gamma(X_t, z, \theta_0)|Z_t = z\right] = 0$ $\forall z \in$Supp(Z_t).

(b) *For all* $\zeta > 0$, $\frac{1}{\sqrt{T}} \sum E\left[|\gamma(X_t, Z_t, \theta_0)| 1\left\{ |\gamma(X_t, Z_t, \theta_0)| > \sqrt{T}\zeta \right\} \middle| Z_t \right] \to_p 0.$

Assumption 6(1) is a standard asymptotic linearity assumption, and the main requirement is that $\hat{\theta}$ be $\sqrt{T}$ consistent, as is usually the case in the stationary parametric context we consider. In particular, if we take $\hat{\theta}$ to be the ML estimator for θ then we can take $\gamma(x, z, \theta_0)$ and I_0 to be the score of the log conditional likelihood and the inverse Fisher Information, respectively, i.e. $\gamma(x, z, \theta_0) = \frac{\partial}{\partial \theta} \log f(x|\theta_0, z)$ and $I_0 = -E\left[\frac{\partial^2}{\partial \theta^2} \log f(X_t|\theta_0, Z_t) \right]^{-1}$. For this choice, sufficient conditions for Assumptions 6(2a) and 6(2b) are that we be able to twice differentiate the log likelihood under an expectation and that the score have $2 + \varepsilon$ moments under the null for $\varepsilon > 0$.

Finally, to derive the limiting distribution of $\sqrt{T} R(\hat{\theta}, \lambda, \xi)$ we need to ensure that the class of functions $\mathcal{R}_T = \{ r_t(\theta, \lambda, \xi) : (\lambda, \xi) \in \Lambda \times \Xi \}$ is well-behaved. Taking

$$C(\xi, \lambda, \xi^*, \lambda^*, \theta, F_Z) = \int \int \begin{bmatrix} r_t(\theta, \lambda, \xi) \\ \gamma(x, z, \theta) \end{bmatrix} \begin{bmatrix} r_t(\theta, \lambda^*, \xi^*) \\ \gamma(x, z, \theta) \end{bmatrix}' dF(x|z, \theta) dF_Z(z) \tag{3}$$

we make the following assumption:

Assumption 7 *Conditional on* $\left\{ Z_t^{(2)} \right\}$, *for any deterministic sequence* $\theta_T \to \theta_0$:

1. *If we define* $\mathcal{R}_T$ *as*

$$\mathcal{R}_T = \{ r_t(\theta_T, \lambda, \xi) | (\lambda, \xi) \in \Lambda \times \Xi \} = \left\{ \left(\frac{g(X_t|\lambda, Z_t)}{f(X_t|\theta_T, Z_t)} - 1 \right) \psi(\Phi(Z_t)'\xi) \middle| (\lambda, \xi) \in \Lambda \times \Xi \right\}$$

   then there exists a random variable H_T *such that* $H_T \geq |r_t(\theta_T, \lambda, \xi)| \ \forall r_t(\theta_T, \lambda, \xi) \in \mathcal{R}_T$ *and, for all* $\zeta > 0$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} E_{\theta_T} \left[|H_T| 1\left\{ |H_T| > \sqrt{T}\zeta \right\} \middle| Z_t \right] \to_p 0.$$

2. *For all* $(x, z) \in Supp(X_t, Z_t)$, $g(x|\lambda, z)$ *is Lipshitz in* λ *with Lipshitz constant* $M_1(x, z)$. *Further, for* $M_2(x, z) = \sup_\lambda |g(x|\lambda, z)|$ *we have*

$$\frac{1}{T} \sum_{t=1}^{T} E_{\theta_T} \left[\left(\frac{M_1(X_t, Z_t) + M_2(X_t, Z_t)}{f(X_t|\theta_T, Z_t)} \right)^2 \middle| Z_t \right] = O_p(1).$$

3. For all (λ, ξ), $(\lambda^*, \xi^*) \in \Lambda \times \Xi$,

$$\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{c} r_t(\theta_T,\lambda,\xi) \\ \gamma(X_t,Z_t,\theta_T) \end{array}\right]\left[\begin{array}{c} r_t(\theta_T,\lambda^*,\xi^*) \\ \gamma(X_t,Z_t,\theta_T) \end{array}\right]' \rightarrow_p C\left(\lambda,\xi,\lambda^*,\xi^*,\theta_0,F_Z\right)$$

where C is as defined in (3) and is nonsingular almost everywhere (in $(\lambda,\xi)\times(\lambda^*,\xi^*)$).

Assumption 7(1) requires that $\mathcal{R}_T$ have a well-behaved envelope function which satisfies a uniform integrability condition. Assumption 7(2) is used to obtain a bound on the entropy of the class $\mathcal{R}_T$ and apply a uniform central limit theorem. In particular, 7(2) imposes a Lipshitz condition on this class, together with a restriction that the Lipshitz constants not be excessively volatile. One unfortunate aspect of this condition is that it rules out $g(x|\lambda, z)$ which are discontinuous in λ: since the condition is sufficient but not necessary for the results that follow, for particular functions of interest that are discontinuous in λ one could likely obtain alternative convergence results. Finally, 7(3) requires the point-wise convergence of the sample covariance function to the limiting covariance function $C\left(\lambda,\xi,\lambda^*,\xi^*,\theta_0,F_Z\right)$. This is needed to obtain finite-dimensional convergence in distribution of $\sqrt{T}R(\hat{\theta},\lambda,\xi)$. Under Assumptions 4-7 we can derive the limiting distribution of $\sqrt{T}R(\hat{\theta},\lambda,\xi)$ under the null of correct specification.

Theorem 3 *Under Assumption 1 and Assumptions 4-7, under H_0 we have that conditional on $\left\{Z_t^{(2)}\right\}$*

$$\sqrt{T}R(\hat{\theta},\lambda,\xi) \Rightarrow \mathbb{G} - \Delta_0(\lambda,\xi)' I_0 \eta_0 = \mathbb{G}^*$$

where $\Rightarrow$ denotes weak convergence (see Van der Vaart and Wellner (1996)) and $\begin{pmatrix} \mathbb{G} \\ \eta_0 \end{pmatrix}$ is a Gaussian process with covariance function $C\left(\xi,\lambda,\xi^*,\lambda^*,\theta_0,F_Z\right)$.

The proof for this result, which can be found in the Appendix, first shows that

$$\begin{pmatrix} \sqrt{T}R(\theta_0,\lambda,\xi) \\ \sqrt{T}(\hat{\theta}-\theta_0) \end{pmatrix} \Rightarrow \begin{pmatrix} \mathbb{G} \\ I_0\eta_0 \end{pmatrix}$$

using Assumptions 1, 4, 6, and 7, and then uses Assumption 5 together with the Continuous Mapping Theorem to obtain the result.

By Theorem 3, we know that the statistic $\sqrt{T}R(\hat{\theta},\lambda,\xi)$ has a well-defined limiting distribution and hence, by the Continuous Mapping Theorem, that under the null of correct specification $||\sqrt{T}R(\hat{\theta},\lambda,\xi)||_\infty \Rightarrow ||\mathbb{G}^*||_\infty$, justifying the use of this statistic as the basis of a test. A problem, however, is that the distribution of $\mathbb{G}^*$ under the null generally depends on θ_0 as well as on the unmodeled distribution of $Z_t^{(2)}$. Hence, while we know that $||\sqrt{T}R(\hat{\theta},\lambda,\xi)||_\infty$ converges to a well-defined limiting distribution under the null, this statistic is not asymptotically pivotal and we must find critical values on a case-by-case basis. To address this challenge we propose a semi-parametric bootstrap approach.

2.2 Bootstrap Critical Values

To calculate critical values we adapt the semi-parametric bootstrap used by Andrews (1997). As noted in that paper, to obtain valid bootstrap critical values we need to impose the null hypothesis in our bootstrap samples, which in our case means imposing that conditional on Z_t, X_t is distributed according to $F(x|Z_t,\theta)$ for some $\theta\in\Theta$. Since we do not model the distribution of $Z_t^{(2)}$ we condition on the realized $\left\{Z_t^{(2)}\right\}_{t=1}^T$ in calculating bootstrap critical values. Formally, our bootstrap algorithm is as follows.

1. Given a sample $\{(X_t,Z_t)\}_{t=1}^T$, calculate $||\sqrt{T}R(\hat{\theta},\lambda,\xi)||_\infty$ as described above and let $b=1$.

2. In bootstrap iteration b, for $t=1,...,T$ draw $X_{t,b}^*$ from $F(x|Z_{t,b}^*,\hat{\theta})$. For each t calculate $Z_{t,b}^{(1)*}$ using $\left(X_{1,b}^*,Z_{1,b}^*\right),...\left(,X_{t-1,b}^*,Z_{t-1,b}^*\right)$ and let $Z_{t,b}^*=(Z_{t,b}^{(1)*},Z_t^{(2)})$ (taking $Z_1^*=Z_1$).

3. Based on bootstrap sample $\left\{(X_{t,b}^*,Z_{t,b}^*)\right\}_{t=1}^T$, calculate estimator $\hat{\theta}_b^*$ for θ and use this to calculate $||\sqrt{T}R_b^*(\hat{\theta}_b^*,\lambda,\xi)||_\infty$ for

$$\sqrt{T}R_b^*(\theta,\lambda,\xi)=\frac{1}{\sqrt{T}}\sum_{t=1}^T\left(\frac{g(X_{t,b}^*|\lambda,Z_{t,b}^*)}{f(X_{t,b}^*|\theta,Z_{t,b}^*)}-1\right)\psi(\Phi(Z_{t,b}^*)'\xi).$$

   Store the resulting value. If $b<B$, let $b=b+1$ and return to step 2.

4. After B iterations, let $c_{\alpha TB}(\hat{\theta})$ be the $1-\alpha$ quantile of $\left\{||\sqrt{T}R_b^*(\hat{\theta}_b^*,\lambda,\xi)||_\infty\right\}_{b=1}^B$. We reject the null hypothesis if and only if $c_{\alpha TB}(\hat{\theta})<||\sqrt{T}R(\hat{\theta},\lambda,\xi)||_\infty$.

To prove the validity of this bootstrap approach we need to show that for $c_\alpha(\theta_0)$ the $1-\alpha$ quantile of $\mathbb{G}^*$ conditional on $\left\{Z_t^{(2)}\right\}$ we have that $c_{\alpha TB}(\hat{\theta}) \rightarrow_p c_\alpha(\theta_0)$ (while the bootstrap algorithm given above uses only $\left\{Z_t^{(2)}\right\}_{t=1}^T$, from the perspective of asymptotic theory we condition on the infinite sequence $\left\{Z_t^{(2)}\right\} = \left\{Z_t^{(2)}\right\}_{t=1}^\infty$). To do this, we show that conditional on $\left\{Z_t^{(2)}\right\}$ for nonrandom sequences $\theta_T \rightarrow \theta_0$ we have that under θ_T, $\sqrt{T}R(\hat{\theta}, \lambda, \xi) \Rightarrow \mathbb{G}^*$, where "under θ_T" means that the data $\{(X_t, Z_t)\}_{t=1}^T$ of sample size T are generated from the DGP with true parameter value θ_T (i.e. $X_t|Z_t \sim F(x|\theta_T, Z_t)$). If we let $c_{\alpha T}(\theta_T)$ be the $1-\alpha$ quantile of $||\sqrt{T}R(\hat{\theta}, \lambda, \xi)||_\infty$ under θ_T for sample size T, the fact that

$$Pr_{\theta_T}\left\{||\sqrt{T}R(\hat{\theta}, \lambda, \xi)||_\infty > c_{\alpha T}(\theta_T)\Big|\left\{Z_t^{(2)}\right\}\right\} = \alpha$$

and $\sqrt{T}R(\hat{\theta}, \lambda, \xi) \Rightarrow \mathbb{G}^*$ implies that, since $||\mathbb{G}^*||_\infty$ has an absolutely continuous distribution by Tsirelson (1976), $Pr_{\theta_T}\left\{||\sqrt{T}R(\hat{\theta}, \lambda, \xi)||_\infty > c_\alpha(\theta_0)\Big|\left\{Z_t^{(2)}\right\}\right\} \rightarrow \alpha$ and hence $c_{\alpha T}(\theta_T) \rightarrow c_\alpha(\theta_0)$ conditional on $\left\{Z_t^{(2)}\right\}$.

The discussion above was all based on the assumption that we had a fixed sequence $\theta_T \rightarrow \theta_0$, while in practice we have only that $\hat{\theta} \rightarrow_p \theta_0$. However, the Almost Sure Representation Theorem (see Theorem 1.10.4 in Van der Vaart and Wellner (1996)) implies that if $\hat{\theta}$ is weakly consistent conditional on $\left\{Z_t^{(2)}\right\}$ we have that $c_{\alpha TB}(\hat{\theta}) \rightarrow_p c_\alpha(\theta_0)$ conditional on $\left\{Z_t^{(2)}\right\}$ provided $B \rightarrow \infty$ as $T \rightarrow \infty$- see also Section 4.1 in Andrews 1997. Summing up, under the null the fact that $c_{\alpha TB}(\hat{\theta}) \rightarrow_p c_\alpha(\theta_0)$ conditional on $\left\{Z_t^{(2)}\right\}$ together with the absolute continuity of $||\mathbb{G}^*||_\infty$ yields that $Pr\left\{||\sqrt{T}R(\hat{\theta}, \lambda, \xi)||_\infty > c_{\alpha TB}(\hat{\theta})\Big|\left\{Z_t^{(2)}\right\}\right\} \rightarrow \alpha$. Provided Assumptions 5-8 hold for all $\theta \in \Theta$, this implies that

$$\sup_{\theta_0 \in \Theta} \lim_{T \rightarrow \infty} Pr_{\theta_0}\left\{||\sqrt{T}R(\hat{\theta}, \lambda, \xi)||_\infty > c_{\alpha TB}(\hat{\theta})\Big|\left\{Z_t^{(2)}\right\}\right\} = \alpha$$

and hence that tests constructed with bootstrap critical values have correct asymptotic size conditional on $\left\{Z_t^{(2)}\right\}$. Interestingly, and importantly for power considerations, if $\hat{\theta} \rightarrow_p \theta_1 \in \Theta$ (and Assumptions 5-8 hold for all $\theta \in \Theta$), the same argument implies that $c_{\alpha TB}(\hat{\theta}) \rightarrow_p c_\alpha(\theta_1)$ conditional on $\left\{Z_t^{(2)}\right\}$ regardless of whether or not the null is true.

To formalize the discussion above, we make the following additional assumption:

Assumption 8 *Conditional on* $\left\{Z_t^{(2)}\right\}$*, for all nonrandom sequences with* $\theta_T \rightarrow \theta_0$*:*

1. *Under* $X_t|\mathcal{F}_{t-1} \sim F(x|Z_t, \theta_T)$, $\sqrt{T}(\hat{\theta} - \theta_T) = \frac{1}{\sqrt{T}} \sum_t I_0 \gamma(X_t, Z_t, \theta_T) + o_p(1)$ *where* I_0 *is a k-dimensional nonrandom matrix which may depend on* θ_0.

2. $\gamma(x, z, \theta)$ *is a measurable function satisfying:*

   (a) *Under* $X_t \sim F(x|z, \theta_T)$, $E[\gamma(X_t, z, \theta_T)|Z_t = z] = 0 \ \forall z \in Supp(Z_t)$.

   (b) *For all* $\zeta > 0$, $\frac{1}{\sqrt{T}} \sum E\left[|\gamma(X_t, Z_t, \theta_T)| 1\left\{ |\gamma(X_t, Z_t, \theta_T)| > \sqrt{T}\zeta \right\} \middle| Z_t \right] \to_p 0$.

This assumption is a straightforward extension of Assumption 6 above and requires that the estimator $\hat{\theta}$ have an linear asymptotic representation under sequences of true parameter values $\theta_T \to \theta_0$. As with Assumption 6, this is not a substantial restriction and generally holds for reasonable estimators. Under these conditions, we obtain the following theorem:

Theorem 4 *Suppose that Assumptions 1, 4, 5, 7, and 8 hold. Then conditional on* $\left\{ Z_t^{(2)} \right\}$ *under* H_0 *under any nonrandom sequence* $\theta_T \to \theta_0$ *we have* $\sqrt{T}R(\hat{\theta}, \lambda, \xi) \Rightarrow \mathbb{G}^*$.

To obtain formal semi-parametric bootstrap results when the null is violated we require that for a given alternative the estimator converge in probability to some value as the sample grows:

Assumption 9 *Conditional on* $\left\{ Z_t^{(2)} \right\}$, $\hat{\theta} \to_p \theta_1 \in \Theta$.

Note that this assumption does not require correct specification. Indeed, results of this sort have been shown to hold in many contexts when the model is misspecified (i.e. when $F_0(\cdot|z) \neq F(\cdot|z, \theta)$ for all $\theta \in \Theta$). For examples and general conditions under which the MLE satisfies Assumption 9, see White (1982). Based on these assumptions, we obtain the following result concerning bootstrap critical values:

Corollary 1 *Denote by* $c_{\alpha TB}(\hat{\theta})$ *the* $1-\alpha$ *quantile of the bootstrap distribution of* $\sqrt{T}R(\hat{\theta}, \lambda, \xi)$, *and by* $c_\alpha(\theta_0)$ *the* $1-\alpha$ *quantile of the distribution of* $||\mathbb{G}^*||_\infty$. *Then conditional on* $\left\{ Z_t^{(2)} \right\}$:

1. *Under Assumptions 1, 4, 5, 7, and 8 and* $B \to \infty$ *as* $T \to \infty$, *we have that under* H_0 $c_{\alpha TB}(\hat{\theta}) \to_p c_\alpha(\theta_0)$ *and* $Pr_{\theta_0}\left(||\sqrt{T}R(\hat{\theta}, \lambda, \xi)||_\infty > c_{\alpha TB}(\hat{\theta}) \middle| \left\{ Z_t^{(2)} \right\} \right) \to \alpha$.

2. *Suppose that Assumptions 1, 4 and 9 hold, that Assumptions 5, 7, and 8 hold at* θ_1, *and that* $B \to \infty$ *as* $T \to \infty$. *Then* $c_{\alpha TB}(\hat{\theta}) \to_p c_\alpha(\theta_1)$.

Corollary 1(1) formalizes the statement that the test with bootstrap critical values has correct asymptotic size conditional on $\left\{Z_t^{(2)}\right\}$, while (2) states that, provided our estimator $\hat{\theta}$ converges to some value θ_1 which satisfies our assumptions, we have that the bootstrap critical value converges to some constant (which may be a function of the true DGP) whether or not the null hypothesis is correct. This reflects the fact that we impose the null of correct specification in our bootstrap samples and plays an important role in the power analysis in the next section.

3 Power of the Proposed Test

Above, we established that tests of $H_0 : X_t|\mathcal{F}_{t-1} \sim F(x|\theta_0, Z_t)$ based on $||\sqrt{T}R(\hat{\theta}, \lambda, \xi)||_\infty$ and bootstrap critical values have correct asymptotic size. What, then, can we say about their power properties? In this section we discuss two results, the first concerning the power of the test against fixed (global) alternatives as the sample size grows and the second concerning the local power of the test.

For questions of both local and global power we will assume that $\hat{\theta} \to_p \theta_1$ for some value θ_1. The key issue in determining whether tests based on $||\sqrt{T}R(\hat{\theta}, \lambda, \xi)||_\infty$ have nontrivial power is the behavior of $\sup_{\xi,\lambda} |E[r_t(\theta_1, \lambda, \xi)]|$. In particular, if $E[r_t(\theta_1, \lambda, \xi)] \equiv 0$, it is clear that tests based on $||\sqrt{T}R(\hat{\theta}, \lambda, \xi)||_\infty$ will not necessarily be consistent. Hence, in our discussion of power we primarily focus on alternatives which violate the tested moment conditions:

Assumption 10 *The true distribution is such that* $\sup_{\xi,\lambda} |E_{f_0}[r_t(\theta_1, \lambda, \xi)]| > 0$.

From Theorems 1 and 2, we know that by choosing $\{G(x|\lambda, z) : \lambda \in \Lambda\}$ to be a complete family we can ensure that Assumption 10 holds whenever $Pr\{f_0(X_t|Z_t) \neq f(X_t|\theta_1, Z_t)\} > 0$, which implies consistency against fixed alternatives which violate H_0^*. In some cases, however, we may not be primarily concerned with consistency and may instead prefer to focus on particular types of deviation from the null. In that case we can choose $\{G(x|\lambda, z) : \lambda \in \Lambda\}$ such that $\sup_{\xi,\lambda} |E_{F_0}[r_t(\theta_1, \lambda, \xi)]|$ is large for true distributions F_0 which deviate from the null in the direction of interest, but $\{G(x|\lambda, z) : \lambda \in \Lambda\}$ may or may not be a complete family. For such a choice Assumption 10 continues to hold for the alternatives of interest and the power results of this section remain valid.

3.1 Power Against Fixed Alternatives

We first consider the global power properties of the test, i.e. the power when the data is drawn from some fixed alternative distribution and the sample size is taken to infinity. To do this, we make the following assumption:

Assumption 11 *Conditional on* $\left\{Z_t^{(2)}\right\}$*, for any sequence* $\theta_T \to \theta_1$ *we have that for all fixed* (λ, ξ)

$$\frac{1}{T}\sum_t r_t(\theta_T, \lambda, \xi) \to_p E\left[r_t(\theta_1, \lambda, \xi)\right].$$

Under this assumption MLR tests are consistent:

Theorem 5 *Under Assumptions 1, 4, 9 and 11, together with Assumptions 5, 7, and 8 (replacing* θ_0 *with* θ_1*), if Assumption 10 holds we have that conditional on* $\left\{Z_t^{(2)}\right\}$

$$\lim_{T\to\infty} Pr\left\{ ||\sqrt{T}R(\hat{\theta}, \lambda, \xi)||_\infty > c_{\alpha TB}(\hat{\theta}) \middle| \left\{Z_t^{(2)}\right\}\right\} = 1.$$

The proof of this result is very straightforward, since by Corollary 1 we have that $c_{\alpha TB}(\hat{\theta}) \to_p c_\alpha(\theta_1) < \infty$, while by Assumption 11 and the Almost Sure Representation Theorem $||\sqrt{T}R(\hat{\theta}, \lambda, \xi)||_\infty \to_p \infty$, delivering the desired result.

By the results of Section 1, we know that if $\{G(x|\lambda, z) : \lambda \in \Lambda\}$ is a complete family we have that $\sup_{\lambda,\xi} |E\left[r_t(\theta_1, \lambda, \xi)\right]| > 0$ whenever H_0^* is violated. Note, however, that the null hypothesis of interest is $H_0 : X_t|\mathcal{F}_{t-1} \sim F(x|\theta_0, Z_t)$, so if we consider alternatives in which X_t depends on $\mathcal{F}_{t-1}$ as a whole rather than just Z_t we can find distributions (in $H_0^*\backslash H_0$) such that $\sup_{\lambda,\xi} |E\left[r_t(\theta_1, \lambda, \xi)\right]| = 0$, against which the MLR test will not necessarily be consistent. This result is entirely intuitive: if we consider an alternative which implies that the distribution of $X_t|Z_t$ is correctly specified but that X_t also depends on some other variable $W_t \in \mathcal{F}_{t-1}$ which is excluded from our test statistic, the MLR test will not in general detect this. Hence, while the MLR test is consistent against general fixed alternatives in independent models (where H_0 and H_0^* are equivalent), in the dependent case it is consistent only against alternatives which imply misspecification in the stationary conditional distribution $X_t|Z_t$ (that is alternatives which violate H_0^*). Nonetheless, if we are interested in a particular collection of alternatives we can always add elements to Z_t to ensure consistency of the MLR test against these alternatives.

3.2 Power Against Local Alternatives

Next, we consider the asymptotic power of the test against local alternatives, i.e. sequences of alternatives which drift closer to the null as the sample size grows. There are two purposes of this exercise: first, it allows us to show that the proposed test has power against alternatives which approach the null at rate $\sqrt{T}$ and, second, it gives some intuition for how the choice of $g(\cdot)$, $\psi(\cdot)$, and $\Phi(\cdot)$ affects the power of the test against different alternatives.

To derive the power of the test against $\sqrt{T}$ local alternatives, we follow the approach of Andrews (1997). In particular, consider a fixed alternative F_0 which has density $f_0(x|z)$ with respect to the dominating measure μ. Let $d(x,z) = \sqrt{T}_0 \left(f_0(x|z) - f(x|z,\theta_0)\right)$ for some value $\theta_0 \in \Theta$ and constant T_0, and consider the sequence of alternative densities

$$f_T(x|z) = f(x|\theta_0, z) + \frac{1}{\sqrt{T}} d(x,z).$$

Note that, by construction, f_T is a valid density for $T \geq T_0$ and $f_{T_0} = f_0$. For the purpose of local power calculations, we consider a sequence of local alternatives where the sample of size T is generated by $X_t|\mathcal{F}_{t-1} \sim F_T(x|Z_t)$ with density $f_T(x|Z_t)$. To ensure that the sequence of local alternatives F_T is contiguous to $F(x|\theta_0, z)$, we make the following assumption:

Assumption 12 *f_0 is such that the following conditions hold conditional on $\left\{Z_t^{(2)}\right\}$:*

1. *Under $F(x|\theta_0,z)$, $\sum_{t=1}^{T} E_{\theta_0}\left[\frac{d(X_t,Z_t)}{f(X_t|\theta_0,Z_t)} 1\left\{\frac{d(X_t,Z_t)}{f(X_t|\theta_0,Z_t)} > \sqrt{T}\zeta\right\}\middle| Z_t\right] \to_p 0 \; \forall \zeta > 0$.*
2. *Under $F(x|\theta_0,z)$, $\frac{1}{T}\sum_{t=1}^{T} \frac{d(X_t,Z_t)^2}{f(X_t|\theta_0,Z_t)^2} \to_p E_{\theta_0}\left[\frac{d(X_t,Z_t)^2}{f(X_t|\theta_0,Z_t)^2}\right]$.*
3. *Under $F(x|\theta_0,z)$, there exists $\delta > 0$ such that*

$$\sup_{0<\varepsilon<\delta}\left\{\frac{1}{T^{\frac{3}{2}}}\sum \frac{|d(X_t,Z_t)|^3}{|f(X_t|\theta_0,Z_t) + \varepsilon d(X_t,Z_t)|^3}\right\} \to_p 0.$$

4. *For all $(\lambda,\xi) \in \Lambda \times \Xi$:*
   (a) *Under F_0, $\frac{1}{T}\sum_{t=1}^{T} E_{f_0}\left[r_t(\theta_0,\lambda,\xi)|Z_t\right] \to_p \int E_{f_0}\left[r_t(\theta_0,\lambda,\xi)|Z_t = z\right] dF_Z$.*
   (b) *Under F_0, $\frac{1}{T}\sum_{t=1}^{T} E_{f_0}\left[\gamma(X_t,Z_t,\theta_0)|Z_t\right] \to_p \int E_{f_0}\left[\gamma(X_t,z,\theta_0)|Z_t = z\right] dF_Z$.*

These conditions require that the distribution f_0 not behave in too extreme a fashion relative to $f(\theta_0)$. In particular, (1) and (2) are what we need in order to apply a martingale central limit theorem to $\frac{d(X_t,Z_t)}{f(X_t|\theta_0,Z_t)}$, while (3) effectively imposes a restriction on the higher moments of this ratio (which is not terribly demanding, since the $T^{-\frac{3}{2}}$ normalization in (3) is quite strong). Together with our earlier assumptions, we obtain the following theorem:

Theorem 6 *Suppose that Assumptions 1 and 4-8 hold at θ_0 and that Assumption 12 is satisfied. Then under $\{F_T(\cdot|\cdot)\}$ we have that conditional on $\left\{Z_t^{(2)}\right\}$*

$$\sqrt{T}R(\hat{\theta},\lambda,\xi) \Rightarrow \mathbb{G}^* + \sqrt{T_0}\nu(\lambda,\xi)$$

where

$$\nu(\lambda,\xi) = \int E_{f_0}\left[r_t(\theta_0,\lambda,\xi)\right]dF_Z - \Delta_0(\lambda,\xi)'I_0\int\int\gamma(x,z,\theta_0)f_0(x|z)d\mu(x)dF_Z.$$

Since one can show that under $\{F_T(\cdot|\cdot)\}$ $\hat{\theta}\to_p\theta_0$, bootstrap critical values will converge to quantiles of $\mathbb{G}^*$ and hence the test will have non-trivial power whenever $\nu(\lambda,\xi)$ is not identically equal to zero.

Note that the limit of $\sqrt{T}R(\hat{\theta},\lambda,\xi)$, and in particular $\nu(\lambda,\xi)$, depends on the value of θ_0, and hence on the point towards which we shrink the sequence of alternatives F_T as the sample size grows. An important question, then, is how we should think about this value θ_0. While the local power results stated above are valid for any value of θ_0 satisfying Assumptions 4-8 and 12, the value of these results (beyond establishing that the test has power against $\sqrt{T}$-local alternatives) is to give us tools for thinking about the power of the test, and in particular thinking about how the power of the test differs across different alternatives. Formally, local power results are useful for this purpose because they allow us to address the case where the sample is "large" (and hence Gaussian approximations to the behavior of $\sqrt{T}R(\hat{\theta},\lambda,\xi)$ are reasonable) without immediately yielding rejection for any violation of the null on the basis of consistency results like Theorem 5. For this purpose, as argued in Andrews (1997) it is natural to let θ_0 be the "quasi-true" value under F_0, since (i) this is the value of θ which makes $F(\theta)$ as "close" as possible the F_0 in the sense of Kullback Leibler divergence, (ii) this choice ensures that $\sqrt{T}(\hat{\theta}-\theta_0)$ has the same behavior under $\{F_T\}$ as under F_{θ_0}, and (iii) for this choice of θ_0, $\hat{\theta}$ has the same probability limit under F_0 and $F(\theta_0)$. Happily, for this choice of θ_0 we have that $\int\int \gamma(x,z,\theta_0) f_0(x|z) d\mu(x) dF_Z = 0$, and hence that ν simplifies to $\int E_{f_0}\left[r_t(\theta_0,\lambda,\xi)|z_t\right] dF_Z$. This means that if $\sup_{\lambda,\xi}\left|E_{f_0}\left[r_t(\theta_0,\lambda,\xi)\right]\right| \neq 0$ (as is always the case if $\{G(x|\lambda) : \lambda \in \Lambda\}$ is a complete family and ψ, Φ satisfy the conditions of Theorem 2) then tests based on $||\sqrt{T}R(\hat{\theta},\lambda,\xi)||_\infty$ have nontrivial power against $\sqrt{T}$ local alternatives of this form.

The term $\int E_{f_0}\left[r_t(\theta_0,\lambda,\xi)|z\right] dF_Z$ can be interpreted as an L^2 inner product. In particular, recall that

$$\int E_{f_0}\left[r_t(\theta_0,\lambda,\xi)|z\right] dF_Z = \int \left\langle \left(\frac{f_0(x|z)}{f(x|\theta_0,z)} - 1\right), g(x|\lambda,z)\psi(\Phi(z)'\xi)\right\rangle dF_Z,$$

the integral over distribution of the conditioning variable of the L^2 inner product of the re-centered likelihood ratio $\left(\frac{f_0(x|z)}{f(x|\theta_0,z)} - 1\right)$ and $g(x|\lambda,z)\psi(\Phi(z)'\xi)$ viewed as functions of x. From this perspective, the test is based on checking orthogonality conditions between the re-centered likelihood ratio $\left(\frac{f_0(x|z)}{f(x|\theta_0,z)} - 1\right)$ (which is identically zero under correct specification) and the continuum of functions $g(x|\lambda,z)$. The test has non-trivial local power against sequences of alternatives such that for some value λ the orthogonality condition

$$\left\langle \left(\frac{f_0(x|Z_t)}{f(x|\theta_0,Z_t)} - 1\right), g(x|\lambda,Z_t)\psi(\Phi(Z_t)'\xi)\right\rangle = 0$$

is violated with positive probability (with respect to F_Z), and the virtue of complete families $\{G(x|\lambda,z) : \lambda \in \Lambda\}$ is that they ensure that such a violation occurs whenever H_0^* is violated. Given this interpretation, if we are interested in a particular alternative $F^*(x|z)$ one reasonable choice for $\{G(x|\lambda,z) : \lambda \in \Lambda\}$ is a complete family that includes $F^*(x|z)$, though many other choices may also make sense.

4 Theoretical Comparison with Previous Literature

The family of tests proposed here and the tests of Andrews (1997) and Bierens and Wang (2012) can all be viewed as extensions of the literature on specification testing using finite-dimensional collections of moment equalities (as in e.g. Newey (1985)) to particular infinite collections chosen to deliver consistent tests (the consistent test of Zheng (2000) is of a somewhat different sort). In particular, Andrews (1997) bases his approach on the statistic

$$A(x,z) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(1\{X_t \le x\} - F(x|\hat{\theta}, Z_t)\Big) 1\{Z_t \le z\}$$

while Bierens and Wang base their test on

$$B(\tau,\xi) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\exp\left(i\tau' X_t\right) - \int \exp\left(i\tau' x\right) dF(x|\hat{\theta}, Z_t)\right) \exp\left(i\xi' Z_t\right).$$

While Andrews considers $||A(x,z)||_\infty$ and Bierens and Wang consider $||B(\tau,\xi)||_2$ (that is, the L^2 norm of B) Andrews notes that, as with the tests proposed here, one could also consider a version of his test based on the L^2 norm so we do not consider this difference to be essential. Instead, the key difference between these statistics and those proposed here is in the continuum of moment equalities being tested. In particular, testing that $E[A(x,z)] \equiv 0$ is the same (asymptotically) as testing that the function $E[(1\{X_t \le x\} - F(x|\theta_1, Z_t)) 1\{Z_t \le z\}]$ is identically equal to zero for $\theta_1 = \text{plim}(\hat{\theta})$. Likewise, testing that $E[B(\tau,\xi)] \equiv 0$ is asymptotically equivalent to testing that $E\left[\left(\exp(i\tau' X_t) - \int \exp(i\tau' x)\, dF(x|\theta_1, Z_t)\right) \exp(i\xi' Z_t)\right] \equiv 0$. The consistency of both tests, and of the MLR tests proposed here, follows from the equivalence of their chosen continuum of moment conditions and the null of correct model specification.

An appealing aspect of the approach taken in this paper (based on testing (2)) relative to those of Andrews (1997) and Bierens and Wang (2012) is that we have a great deal of freedom in choosing g, ψ and Φ, so if we are concerned with alternatives that imply particular behavior for the likelihood ratio $\frac{f_0(X_t|Z_t)}{f(X_t|\theta_1,Z_t)} - 1$ we can construct tests that have higher power in those directions while maintaining consistency against global alternatives. Moreover, the class of functions $\left\{\left(\frac{g(X_t|\lambda,Z_t)}{f(X_t|\theta_0,Z_t)} - 1\right)\psi(\Phi(Z_t)'\xi) : (\lambda,\xi) \in \Lambda \times \Xi\right\}$ are martingale differences under the null of correct specification, greatly simplifying the problem of testing under dependence and allowing treatment of the independent and dependent cases in a unified manner.[2]

[2]Note that the statistics considered by Andrews (1997) and Bierens and Wang (2012) also result in collections martingales under the null of correct specification, and hence it may be possible to extend their results to dependent models by arguments analogous to those in the present paper, though we do not pursue such extensions here.

4.1 Relation to Bartlett Identities Tests

Certain MLR tests are closely related to the Information Matrix test proposed by White (1982) and the Bartlett Identities tests proposed by Chesher et. al. (1999). In particular Chesher et. al. consider generalizations of White's Information Matrix test based on the observation that if the true conditional density is $f(x|\theta_0, z)$, then so long as we can exchange integration and differentiation of the log likelihood k times we have that $E\left[\frac{\partial^k f(X_t|\theta_0,Z_t)/\partial\theta^k}{f(X_t|\theta_0,Z_t)}\right] = 0$ (for ease of exposition we will treat θ as scalar, but all results extend directly to the general case). To exploit this relationship, they propose tests based on finite collections of statistics $\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^k f(X_t|\hat{\theta},Z_t)/\partial\theta^k}{f(X_t|\hat{\theta},X_t)}$, while White's Information Matrix test considers these statistics for k=2. Note, however, that if the density $f(x|\theta, z)$ admits a Taylor series representation on some open neighborhood $\tilde{\Theta}$ of θ_0 then for $\theta \in \tilde{\Theta}$ we have $f(x|\theta, z) = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{\partial^k f(X_t|\hat{\theta},Z_t)}{\partial\theta^k} (\theta - \theta_0)^k$ and if we can exchange an expectation with the infinite sum we have that the restriction $E\left[\frac{\partial^k f(X_t|\theta_0,Z_t)/\partial\theta^k}{f(X_t|\theta_0,Z_t)}\right] = 0 \ \forall k \in \mathbb{N}$ is equivalent to

$$E\left[\sum_{k=0}^{\infty} \frac{1}{k!} \frac{\partial^k f(X_t|\theta_0, Z_t)/\partial\theta^k}{f(X_t|\theta_0, Z_t)} (\theta - \theta_0)^k\right] = E\left[\frac{f(X_t|\theta, Z_t)}{f(X_t|\theta_0, Z_t)}\right] = 1 \quad \forall\theta \in \tilde{\Theta}.$$

Note, however, that the restriction $E\left[\frac{f(X_t|\theta,Z_t)}{f(X_t|\theta_0,Z_t)} - 1\middle| Z_t\right] = 0$ $a.s.$ which is tested by the MLR test with $\{G(x|\lambda, z)|\lambda \in \Lambda\} = \left\{F(x|\lambda, z)|\lambda \in \tilde{\Theta}\right\}$ (i.e. taking g to be a density from the same parametric family assumed to generate the data) implies the restriction $E\left[\frac{f(X_t|\theta,Z_t)}{f(X_t|\theta_0,Z_t)}\right] = 1$ $\forall\theta \in \tilde{\Theta}$. Hence, MLR tests based on this choice of G (setting $\psi(\cdot) \equiv 1$) can be viewed as tests of all the Bartlett Identities jointly. Alternatively, the Bartlett Identities can also be shown to hold conditionally, $E\left[\frac{\partial^k f(X_t|\theta_0,Z_t)/\partial\theta^k}{f(X_t|\theta_0,Z_t)}|Z_t\right] = 0$ $a.s.$, and MLR tests with $\{G(x|\lambda, z)|\lambda \in \Lambda\} = \left\{F(x|\lambda, z)|\lambda \in \tilde{\Theta}\right\}$ and ψ chosen to yield consistency can be viewed as tests of this stronger set of restrictions. As discussed by Chesher et. al., tests of the Bartlett Identities can be viewed as tests for unobserved parameter heterogeneity, extending the interpretation suggested by Chesher (1984) for White's Information Matrix test. In cases where the above assumptions regarding Taylor expansion and interchange of differentiation and integration hold, this interpretation applies to MLR tests for this choice of G as well.

5 Simulation Results

In this section, we provide simulation evidence on the finite-sample size and power of particular MLR tests relative to the tests proposed by Andrews (1997), Zheng (2000), and Bierens and Wang (2012). For ease of comparison, in most of our simulations we follow the simulation design of Bierens and Wang (2012) and reproduce some of their tables. Unlike the other tests considered we've proved that the MLR tests are also applicable to dependent data, which we illustrate with an application to an AR(1) model. In each simulation design we consider a number of different choices for $g(x|\lambda, z)$, demonstrating that different choices yield tests with power in different directions.

5.1 Independent Case

5.1.1 Linear Regression Model

Following the simulation design of Zheng (2000) and Bierens and Wang (2012) we begin by considering a linear regression model with $X_t = Z_{1,t} + Z_{2,t} + U_t$, where $Z_{1,t} \equiv 1$ and $Z_{2,t} \sim N(0,1)$, and consider a number of of different distributions for the error term U_t:

$$\begin{gathered} H_Z^{(0)} : U_t|Z_t \sim N(0,1) \\ H_Z^{(1)} : U_t|Z_t \sim \text{Standard Logistic} \\ H_Z^{(2)} : U_t|Z_t \sim t_5 \\ H_Z^{(3)} : U_t|Z_t \sim N(0, Z_t^2) \end{gathered} \tag{4}$$

For each true error distribution, we test the null hypothesis $H_0 : X_t = Z_t\beta + U_t$ for $U_t|Z_t \sim N(0, \sigma^2)$. Again following Bierens and Wang (2012), we simulate the performance of our proposed test based on 1000 samples of size 200 and 500 bootstrap replications, taking $\psi(\cdot) = \exp(\cdot)$ and $\Phi(z) = (F_N(z_1), F_N(z_2))'$ for F_N the normal CDF. We consider five different choices for $g(x|\lambda, z)$, yielding five different MLR tests which we label $\text{MLR}_Z^{(N)}$ and $\text{MLR}_Z^{(0)} - \text{MLR}_Z^{(3)}$. $\text{MLR}_Z^{(N)}$ takes $g(x|\lambda, z) = \phi(x - \lambda)$ where ϕ is a standard normal pdf, while for $i \in \{0, 1, 2, 3\}$, $\text{MLR}_Z^{(i)}$ takes $g(x|\lambda, z) = f(x - z_2 - \lambda)$ where $f(u)$ is the density of U corresponding to $H^{(i)}$. For example $\text{MLR}_Z^{(0)}$ takes $g(x|\lambda, z) = \phi(x - z_2 - \lambda)$. In all cases we set $\Lambda = [-2, 2]$ and $\Xi = [-2, 2]^2$. We consider tests with nominal size 1%, 5%, and 10% and compare the performance of our proposed procedures to that of the tests reported in Table 1 in Bierens and Wang (2012). Table 1 reports the simulation results for the MLR tests together with results (from Bierens and Wang (2012)) for the CK test (1) from Andrews (1997), Zheng's test for three different bandwidth choices (0.5, 1.0, and 2.0), and the Bierens and Wang tests for c=5 (generally among the cases with the highest power) and their Max test.

These results show that all of the MLR tests considered control size reasonably well, though many of the tests are somewhat undersized. For $i \in \{1,2,3\}$ $\text{MLR}_Z^{(i)}$ has reasonably good power against alternative $H_Z^{(i)}$ confirming that, as suggested at the end of Section 3, if we're interested in tests with good power against a particular alternative $F^*(x|z)$ one option is to choose $g(x|\lambda, z)$ so that $\{G(x|\lambda, z) : \lambda \in \Lambda\}$ includes $F^*(x|z)$. Relative to previously proposed tests the MLR tests perform fairly well: against $H^{(1)}$ and $H^{(2)}$, the most powerful MLR test (in both cases $\text{MLR}_Z^{(1)}$) has power competitive with most powerful alternative test (in both cases Z(2.0)). Against $H^{(3)}$ the MLR tests are less powerful than the tests proposed by Andrews and Bierens and Wang but the $\text{MLR}_Z^{(1)}$ test still does reasonably well.

5.1.2 Poisson Regression Model

We next consider the discrete simulation example discussed in Bierens and Wang (2012). In that example, we test the null that $H_0 : X_t|Z_t \sim \text{Poisson}\left(\exp\left(Z_t\beta\right)\right)$ where we again have $Z_{1,t} \equiv 1$, $Z_{2,t} \sim N(0,1)$ and the true conditional DGP is one of:

$$\begin{aligned} H_P^{(0)} &: X_t|Z_t \sim \text{Poisson}\left(\exp\left(Z_{2,t}\right)\right) \\ H_P^{(1)} &: X_t|Z_t \sim \text{NB}\left(1, p(Z_{2,t})\right) \\ H_P^{(2)} &: X_t|Z_t \sim \text{NB}\left(5, p(Z_{2,t})\right) \\ H_P^{(3)} &: X_t|Z_t \sim \text{NB}\left(10, p(Z_{2,t})\right) \end{aligned} \tag{5}$$

where NB denotes the negative binomial distribution and $p(z_2) = \frac{1}{1+e^{-z_2}}$ is the Logit function. Note that the mean of an $\text{NB}\left(r, p(z_2)\right)$ distribution is re^{z_2} while the mean of a $\text{Poisson}\left(\exp\left(z\beta\right)\right)$ is $e^{z_1\beta_1}e^{z_2\beta_2}$, so all of the alternatives imply that the conditional mean of $X_t|Z_t$ is correctly specified (i.e. consistent with H_0). To construct mean likelihood ratio tests we again take $\psi(\cdot) = \exp(\cdot)$ and $\Phi(z) = (F_N(z_1), F_N(z_2))'$. As in the linear regression case we consider five different choices for $g(x|\lambda, z)$ and label the resulting tests $\text{MLR}_P^{(N)}$ and $\text{MLR}_P^{(0)} - \text{MLR}_P^{(3)}$.

$\text{MLR}_P^{(N)}$ and $\text{MLR}_P^{(0)}$ take $g(x|\lambda, z)$ to be the probability mass function for a Poisson(exp(λ)) and a Poisson $(\exp(z_2 \cdot \lambda))$ distribution, respectively, while for $i \in \{1,2,3\}$ $\text{MLR}_P^{(i)}$ takes $g(x|\lambda, z)$ to be the probability mass function for a NB $(k_i, p(z_2 + \lambda))$ distribution, where k_i is as in $H_p^{(i)}$. We again take $\Lambda = [-2, 2]$ and $\Xi = [-2, 2]^2$ in all cases. Results from these simulations are reported in Table 2, where the results for the non-MLR tests are taken from Table 2 in Bierens and Wang (2012).

We can see that the MLR tests again control size reasonably well, though $MLR_P^{(3)}$ appears to be slightly oversized. For $i \in \{1,2,3\}$ we again have that the test $\text{MLR}_P^{(i)}$ provides good power against alternative $H_P^{(i)}$, and in all cases the MLR_P tests have power which is quite competitive with that of existing procedures. In particular, against $H_p^{(1)}$ the $\text{MLR}_P^{(1)} - \text{MLR}_P^{(3)}$ tests have power close to that of the most powerful test considered (the CK test of Andrews 1997), while against $H_P^{(2)}$ and $H_P^{(3)}$ the power of the $\text{MLR}_P^{(1)} - \text{MLR}_P^{(3)}$ tests everywhere exceeds that of the competing tests, often by a substantial margin.

5.2 Dependent Case

Unlike the other consistent tests available in the literature, mean likelihood ratio tests can also be applied in contexts with dependent data. To illustrate this point, we consider an AR(1) variant of the regression model considered above. In particular, we simulate tests of the null $H_0 : X_t = \alpha + \beta X_{t-1} + U_t$ for $U_t \sim N(0, \sigma^2)$ when the true DGP is $X_t = 1 + .7 \cdot X_{t-1} + U_t$ and $U_t|X_{t-1}$ has one of the following distributions:

$$\begin{gathered} H_{AR}^{(0)} : U_t|X_{t-1} \sim N(0,1) \\ H_{AR}^{(1)} : U_t|X_{t-1} \sim \text{Standard Logistic} \\ H_{AR}^{(2)} : U_t|X_{t-1} \sim t_5 \\ H_{AR}^{(3)} : U|X_{t-1} \sim N(0, X_{t-1}^2) \end{gathered} \tag{6}$$

We take $Z_{1,t} \equiv 1$, $Z_{2,t} = X_{t-1}$, $\psi(\cdot) = \exp(\cdot)$ and $\Phi(z) = (F_N(z_1), F_N(z_2))'$. As before we consider a number of choices for $g(x|\lambda, z)$, yielding tests $\text{MLR}_{AR}^{(N)}$ and $\text{MLR}_{AR}^{(0)} - \text{MLR}_{AR}^{(3)}$. $\text{MLR}_{AR}^{(N)}$ takes $g(x|\lambda, z) = \phi(x - \lambda)$, while for $i \in \{0,1,2,3\}$ $\text{MLR}_{AR}^{(i)}$ takes $g(x|\lambda, z) = f^{(i)}(x - .7 \cdot z_2 - \lambda)$ where $f^{(i)}(u)$ is the density of U_t implied by $H_{AR}^{(i)}$. Results from this simulation are reported in Table 3. From these results, we again see that for $i \in \{1,2,3\}$ the

$MLR^{(i)}$ test, while not always the most powerful test considered, has reasonably good power against $H_{AR}^{(i)}$.

6 Conclusion

In this paper we introduce a novel class of specification tests for general parametric models. We show that the hypothesis of correct specification is equivalent to a continuum of moment conditions, and use this to prove that the proposed tests are consistent against fixed alternatives in independent models and against alternatives which depend on the past only through a known finite vector of state variables in dependent models. We further show that these tests are have power against $\sqrt{T}$ local alternatives. In simulation, we show that these tests control size well in finite samples and have power competitive with existing tests in independent models while also performing well in a dependent example. The family of proposed tests is quite broad, and as we illustrate one can choose different tests from this family to increase power against particular alternatives when desired.

Appendix

Proof of Theorem 1 Proof given in text preceding Theorem 1.□

Proof of Theorem 2 Note that since Φ is bounded and Ξ is compact, the existence of

$$E_{f_0}\left[\left(\frac{g(X_t|\lambda, Z_t)}{f(X_t|\theta_0, Z_t)} - 1\right)\psi(\Phi(Z_t)'\xi)\right]$$

follows immediately from the assumption that $E_{f_0}\left[\frac{g(X_t|\lambda,Z)}{f(X_t|\theta_0,Z_t)} - 1\right]$ exists. One direction of the result is immediate: if $f_0(X_t|Z_t) = f(X_t|\theta_0, Z_t)$ almost surely with respect to the true probability measure F_{XZ}, then $E_{f_0}\left[\frac{g(X_t|\lambda,Z_t)}{f(X_t|\theta_0,Z_t)} - 1\middle| Z_t\right] = 0$ almost surely. Hence, (2) holds.

To obtain the converse, suppose that $f_0(X_t|Z_t)$ and $f(X_t|\theta_0, Z_t)$ differ with positive probability (with respect to F_{XZ}). Then there exists a set A such that $Pr\{Z_t \in A\} > 0$ and for $z \in A$, $f_0(x|z) \neq f(x|\theta_0, z)$. By the result of Theorem 1, we know that for each such $z \in A$ there exists $\lambda(z)$ such that $E_{f_0}\left[\frac{g(X_t|\lambda(z),z)}{f(X_t|\theta_0,z)} - 1\middle| Z_t = z\right] \neq 0$.

Consider first the case in which there exists some z^* such that $Pr\{Z_t = z^*\} > 0$ and $E_{f_0}\left[\frac{g(X_t|\lambda(z^*),z^*)}{f(X_t|\theta_0,z^*)} - 1\middle| Z_t = z^*\right] \neq 0$. Since z^* is observed with positive probability, by Theorem 2.3 of Stinchcombe and White (1997) the set of ξ such that $E_{f_0}\left[\left(\frac{g(X_t|\lambda^*,Z_t)}{f(X_t|\theta_0,Z_t)} - 1\right)\psi(\Phi(Z_t)'\xi)\right] = 0$ is of Lebesgue measure zero (and is not dense in $\mathbb{R}^l$), so there exists $\xi^* \in \Xi$ such that $E_{f_0}\left[\left(\frac{g(X_t|\lambda(z^*),Z_t)}{f(X_t|\theta_0,Z_t)} - 1\right)\psi(\Phi(Z_t)'\xi^*)\right] \neq 0$, proving the result.

Consider next the case in which no such z^* occurs with positive probability. Nonetheless, there exists some $\lambda^* \in \Lambda$ such that $Pr\left\{E_{f_0}\left[\frac{g(X_t|\lambda^*,Z_t)}{f(X_t|\theta_0,Z_t)} - 1\middle| Z_t\right] \neq 0\right\} > 0$. To see that this is the case, note that since $E_{f_0}\left[\frac{g(x|\lambda,Z_t)}{f(x|\theta_0,Z_t)} - 1\middle| Z_t\right]$ is continuous in λ almost surely by Assumption 3, Assumption 2 implies that for almost every $z \in A$ the set

$$B(z) = \left\{\lambda : E_{f_0}\left[\frac{g(x|\lambda(z),z)}{f(x|\theta_0,z)} - 1\middle| Z_t = z\right] \neq 0\right\}$$

contains a nonempty open set. To prove the theorem it suffices to show that there exists $\lambda^* \in \Lambda$ such that $Pr\{\lambda^* \in B(Z_t)\} > 0$. Let D be a countable dense subset of Λ, and note that the event $\{D \cap \text{int}(B(z)) \neq \emptyset\}$ (where $\text{int}(B)$ denotes the interior of the set B) is equivalent to the event $\text{int}(B(z)) \neq \emptyset$. Moreover, $D \cap \text{int}(B(z)) \neq \emptyset$ implies that $\lambda \in B(z)$ for some $\lambda \in D$. By countable additivity,

$$Pr\{Z_t \in A\} = Pr\{D \cap \text{int}(B(Z_t)) \neq \emptyset\} \leq \sum_{\lambda \in D} Pr\{\lambda \cap \text{int}(B(Z_t)) \neq \emptyset\}.$$

Hence, since we have that $Pr\{\text{int}(B(Z_t)) \neq \emptyset\} > 0$, it must be the case that there is some $\lambda^* \in D$ such that

$$Pr\{\lambda^* \cap \text{int}(B(Z_t)) \neq \emptyset\} = Pr\left\{E_{f_0}\left[\frac{g(X_t|\lambda^*,Z_t)}{f(X_t|\theta_0,Z_t)} - 1\middle| Z_t\right] \neq 0\right\} > 0,$$

as we wanted to show.

Now that we've established that for some λ^*, $Pr\left\{E_{f_0}\left[\frac{g(X_t|\lambda^*,Z_t)}{f(X_t|\theta_0,Z_t)} - 1\middle| Z_t\right] \neq 0\right\} > 0$, we can apply Theorem 2.3 of Stinchcombe and White (1997), obtaining that

$$E_{f_0}\left[\left(\frac{g(X_t|\lambda^*,Z_t)}{f(X_t|\theta_0,Z_t)} - 1\right)\psi(\Phi(Z_t)'\xi)\right] \neq 0$$

except for a set of ξ of Lebesgue measure zero. This concludes the proof.□

Proof of Theorem 3: This is a special case of Theorem 4, with Assumption 6 everywhere replacing Assumption 8.□

Proof of Theorem 4: By Lemma A.1 we have that for $\gamma_T = \frac{1}{T}\sum \gamma(X_t, Z_t, \theta_T)$

$$\sqrt{T} \sup_{\lambda,\xi} |R(\hat{\theta}, \lambda, \xi) - R(\theta_T, \lambda, \xi) + \Delta_0(\lambda, \xi)' I_0 \gamma_T| = o_p(1)$$

while by Lemma A.4 we have

$$\begin{bmatrix} \sqrt{T} R(\theta_T, \lambda, \xi) \\ \sqrt{T}\gamma_T \end{bmatrix} \Rightarrow \begin{pmatrix} \mathbb{G}_1 \\ \eta_0 \end{pmatrix}.$$

Hence, the conclusion follows by the Continuous Mapping Theorem.□

Lemma A.1 Under the assumptions of Theorem 4, we have that under θ_T

$$\sqrt{T} \sup_{\lambda,\xi} |R(\hat{\theta}, \lambda, \xi) - R(\theta_T, \lambda, \xi) + \Delta_0(\lambda, \xi)' I_0 \gamma_T| = o_p(1).$$

Proof: Note that under Assumption 8 we have that $\sqrt{T}(\hat{\theta} - \theta_T) = O_p(1)$ under θ_T. Next, by Assumption 5 we can use a mean-value expansion

$$\sqrt{T} R(\hat{\theta}, \lambda, \xi) = \sqrt{T} R(\theta_T, \lambda, \xi) - \frac{1}{T}\sum_{t=1}^{T} \frac{\partial}{\partial\theta} f(\theta^*(\lambda, \xi)) \frac{g(X_t|\lambda, Z_t)}{f^2(\theta^*(\lambda, \xi))} \psi(\Phi(Z_t)'\xi) \sqrt{T}(\hat{\theta} - \theta_T)$$

where $\theta^*(\lambda, \xi)$ lies between $\hat{\theta}$ and θ_T. However, Assumption 8 implies that $\hat{\theta} - \theta_0 = o_p(1)$, and hence that there exists a sequence of constants $\delta_T^* \to 0$ such that $Pr\left\{\sup_{\lambda,\xi} ||\theta^*(\lambda, \xi) - \theta_0|| > \delta_T^*\right\} \to 0$. Define

$$N_{1,T} = \sup_{\xi,\lambda} |\frac{1}{T}\sum \frac{\partial}{\partial\theta} f(\theta^*(\lambda, \xi)) \frac{g(X_t|\lambda, Z_t)}{f^2(\theta^*(\lambda, \xi))} \psi(\Phi(Z_t)'\xi) - \Delta_0(\lambda, \xi)|$$

$$N_{2,T} = \sup_{\xi,\lambda} \sup_{\theta:||\theta-\theta_0||\le\delta_T^*} |\frac{1}{T}\sum \frac{\partial}{\partial\theta} f(\theta) \frac{g(X_t|\lambda, Z_t)}{f^2(\theta)} \psi(\Phi(Z_t)'\xi) - \Delta_0(\lambda, \xi)|.$$

By construction, $Pr\left\{N_{2,T} \geq N_{1,T}\right\} \to 1$, but we know by Assumption 5 that $N_{2,T} \to 0$, so $N_{1,T} = o_p(1)$. Note, however, that

$$\sqrt{T}\sup_{\lambda,\xi}|R(\hat{\theta},\lambda,\xi) - R(\theta_T,\lambda,\xi) + \Delta_0(\lambda,\xi)' I_0 \gamma_T|$$

$$= \sqrt{T}\sup_{\xi,\lambda}|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\theta}f(\theta^*(\lambda,\xi))\frac{g(X_t|\lambda,Z_t)}{f^2(\theta^*(\lambda,\xi))}\psi(\Phi(Z_t)'\xi)(\hat{\theta}-\theta_T) - \Delta_0(\lambda,\xi)' I_0\gamma_T|$$

$$\leq N_{1,T}\sqrt{T}||\hat{\theta}-\theta_T|| + \sup_{\lambda,\xi}|\Delta_0(\lambda,\xi)'\left(\sqrt{T}(\hat{\theta}-\theta_T) - \sqrt{T}I_0\gamma_T\right)| = o_p(1)$$

where the final term is $o_p(1)$ by Assumptions 5(3) and 8(1).□

Lemma A.2 Assumption 7(2) implies that $\frac{g(x|\lambda,z)}{f(x|\theta,z)}\psi(\Phi(z)'\xi)$ is Lipshitz in (λ,ξ) with Lipshitz constants $K(x,z,\theta)$ which satisfy $\frac{1}{T}\sum_{t=1}^{\infty}E\left[K(X_t,Z_t,\theta_T)^2|Z_t\right] = O_p(1)$.

Proof: From Assumption 7(2) we have that $\frac{g(x|\lambda,z)}{f(x|\theta,z)}\psi(\Phi(z)'\xi)$ is Lipshitz in λ with Lipshitz constant bounded by $\frac{M_1(x,z)}{f(x|z,\theta)}B_1$ for some constant B_1 (using the boundedness of $\psi(\Phi(z)'\xi)$). Further, since Φ is a bounded function and ψ is continuously differentiable we have that $\psi\left(\Phi(z_t)'\xi\right)$ is Lipshitz in ξ with a Lipshitz constant bounded by some B_2, so $\frac{g(x|\lambda,z)}{f(x|\theta,z)}\psi(\Phi(z)'\xi)$ is Lipshitz in ξ with Lipshitz constant bounded by $\frac{M_2(x,z)}{f(x|\theta,z)}B_2$. Combining these results and applying the triangle inequality, we have that $\frac{g(x|\lambda,z)}{f(x|\theta,z)}\psi(\Phi(z)'\xi)$ is Lipshitz in (λ,ξ) with Lipshitz constant bounded above by $K(x,z,\theta) = \frac{M_1(x,z)}{f(x|\theta,z)}B_1 + \frac{M_2(x,z)}{f(x|\theta,z)}B_2$. Assumption 7(2) gives us that $\frac{1}{T}\sum_{t=1}^{T}E_{\theta_T}\left[\left(\frac{M_1(x,Z_t)+M_2(x,Z_t)}{f(X_t|\theta_T,Z_t)}\right)^2\middle| Z_t\right] = O_p(1)$, which implies the result.□

Lemma A.3 Under Assumption 4 and $H_0: X_t|\mathcal{F}_{t-1} \sim F(x|\theta,Z_t)$ we have that for all $(\lambda,\xi) \in \Lambda\times\Xi$, $\sum_{t=1}^{T}\left(\frac{g(X_t|\lambda,Z_t)}{f(X_t|\theta,Z_t)} - 1\right)\psi(\Phi(Z_t)'\xi)$ is a martingale with respect to $\{\mathcal{F}_t\}_{t=1}^{T}$ conditional on $\left\{Z_t^{(2)}\right\}$.

Proof: As argued in the text, we know that $E_\theta\left[\left(\frac{g(X_t|\lambda,Z_t)}{f(X_t|\theta,Z_t)} - 1\right)\psi(\Phi(Z_t)'\xi)\middle| Z_t\right] = 0$. Next, note that by Assumption 4:

$$E\left[\left(\frac{g(X_t|\lambda,Z_t)}{f(X_t|\theta,Z_t)} - 1\right)\psi(\Phi(Z_t)'\xi)\middle|\mathcal{F}_{t-1},\left\{Z_t^{(2)}\right\}\right] = E\left[\left(\frac{g(X_t|\lambda,Z_t)}{f(X_t|\theta,Z_t)} - 1\right)\psi(\Phi(Z_t)'\xi)\middle|\mathcal{F}_{t-1}\right]$$

since conditional on the values of $Z_t^{(2)}$ up to time t, X_t and all future values of $Z_t^{(2)}$ are independent. However, under H_0 the distribution of X_t depends on $\mathcal{F}_{t-1}$ only through Z_t and hence

$$E\left[\left(\frac{g(X_t|\lambda, Z_t)}{f(X_t|\theta, Z_t)} - 1\right)\psi(\Phi(Z_t)'\xi)\middle|\mathcal{F}_{t-1}\right] = E\left[\left(\frac{g(X_t|\lambda, Z_t)}{f(X_t|\theta, Z_t)} - 1\right)\psi(\Phi(Z_t)'\xi)\middle| Z_t\right] = 0$$

which proves the result.□

Lemma A.4 Under Assumptions 1, 4, 7, 8, and $\theta_T \to \theta_0$ we have that under H_0 :

$$\begin{bmatrix} \sqrt{T}R(\theta_T, \lambda, \xi) \\ \sqrt{T}\gamma_T \end{bmatrix} \Rightarrow \begin{pmatrix} \mathbb{G}_1 \\ \eta_0 \end{pmatrix}$$

in the usual sense of weak convergence.

Proof: For any finite collection collection of elements (λ_1, ξ_1),...(λ_k, ξ_k) of $\Lambda \times \Xi$, we know that the elements of

$$\begin{pmatrix} \sqrt{T}R(\theta_T, \lambda_1, \xi_1) \\ \vdots \\ \sqrt{T}R(\theta_T, \lambda_k, \xi_k) \\ \sqrt{T}\gamma_T \end{pmatrix} \tag{7}$$

satisfy a Lindeberg condition under Assumptions 7(1) and 8(2b) and hence, by repeated application of the triangle inequality, that they satisfy a joint Lindeberg condition as well. By Assumption 7(3), we know that the sample covariances $\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} r_t(\theta_T, \lambda, \xi) \\ \gamma(X_t, Z_t, \theta_T) \end{bmatrix}\begin{bmatrix} r_t(\theta_T, \lambda', \xi') \\ \gamma(X_t, Z_t, \theta) \end{bmatrix}'$ converge to C as in (1). By Lemma A.3 we know that $T(R(\theta_T, \lambda, \xi), \gamma_T)$ is a martingale with respect to $\{\mathcal{F}_t\}$, so by the Martingale Central Limit Theorem for triangular arrays (see e.g. Lipster and Shirayev, Chapter 5, Theorem 8) we have that all finite-dimensional vectors of the form (7) converge to normal distributions with covariance matrix given by C. To obtain functional convergence we combine this finite-dimensional result with Corollary 4.3 in Nishiyama (2000). In particular, note that Assumption 7(1) implies Nishiyama's Assumption [L1']. Further, by Lemma A.2, $\mathcal{R}_T$ consists of a set of Lipshitz functions indexed by $\Lambda \times \Xi$ with Lipshitz constant K satisfying $\frac{1}{T}\sum_{t=1}^{\infty} E\left[K(X_t, Z_t, \theta_T)^2|Z_t\right] = O_p(1)$. Note, however, that $\Lambda \times \Xi$ is a compact subset of Euclidean space, and hence for ρ the usual Euclidean metric trivially satisfies entropy condition $\int_0^1 \sqrt{N(\Lambda \times \Xi, \rho, \varepsilon)} d\varepsilon < \infty$. Hence, by Proposition 4.5 of Nishiyama (2000), condition [PE'] holds and we can apply Corollary 4.3 to obtain the desired result.□

Proof of Corollary 1: Given in text preceding Assumption 9.

Proof of Theorem 5: By Assumptions 9 and 11, we have that $\sup_{(\lambda,\xi)} ||\sqrt{T} R(\hat{\theta}, \lambda, \xi)||_\infty \to_p \infty$ by the Almost-Sure Representation Theorem (Theorem 1.10.4 in Van der Vaart and Wellner, (1996)), which together with Corollary 1 proves the result.

Proof of Theorem 6: As a first step, we need to prove that, conditional on $\left\{Z_t^{(2)}\right\}$, the distribution of $\{(X_t, Z_t) : t \leq T\}$ where the X_t's have conditional distribution $F_T(\cdot|Z_t)$, is contiguous to the distribution of $\{(X_t, Z_t) : t \leq T\}$ where the X_t's are have conditional distribution $F(\cdot|Z_t, \theta_0)$. To do so, consider the log-likelihood ratio $\log\left(\prod_{t=1}^T \frac{f(X_t|\theta_0, Z_t) + \frac{1}{\sqrt{T}} d(X_t, Z_t)}{f(X_t|\theta_0, Z_t)}\right)$. Following Andrews (1997), third-order Taylor expansion of $\log\left(f(X_t|\theta_0, Z_t) + \frac{1}{\sqrt{T}} d(X_t, Z_t)\right)$ around $f(X_t|\theta_0, Z_t)$ yields that with probability tending to one

$$\left|\log\left(\prod_{t=1}^T \frac{f(X_t|\theta_0, Z_t) + \frac{1}{\sqrt{T}} d(X_t, Z_t)}{f(X_t|\theta_0, Z_t)}\right) - \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{d(X_t, Z_t)}{f(X_t|\theta_0, Z_t)} + \frac{1}{2T} \sum_{t=1}^T \frac{d(X_t, Z_t)^2}{f(X_t|\theta_0, Z_t)^2}\right|$$

$$\leq \frac{1}{6} \sup_{0<\varepsilon<\delta} \left\{\frac{1}{T^{\frac{3}{2}}} \sum \frac{|d(X_t, Z_t)|^3}{|f(X_t|\theta_0, Z_t) + \varepsilon d(X_t, Z_t)|^3}\right\}.$$

Note, further, that by Assumption 12(3) the last term is $o_p(1)$ under $F(x|\theta_0, z)$. As a result the limiting behavior of the log likelihood ratio is determined by the first two terms of its Taylor expansion. Note, however, that under $F(x|\theta_0, z)$ $\frac{d(X_t, Z_t)}{f(X_t|\theta_0, Z_t)}$ is a martingale difference. Hence, using Assumption 12 parts (1) and (2) we can apply the martingale central limit theorem and obtain that under $F(x|\theta_0, z)$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{d(X_t, Z_t)}{f(X_t|\theta_0, Z_t)} - \frac{1}{2T} \sum_{t=1}^T \frac{d(X_t, Z_t)^2}{f(X_t|\theta_0, Z_t)^2} \to_d N\left(-\frac{1}{2} E_{\theta_0}\left[\frac{d(X_t, Z_t)^2}{f(X_t|\theta_0, Z_t)^2}\right], E_{\theta_0}\left[\frac{d(X_t, Z_t)^2}{f(X_t|\theta_0, Z_t)^2}\right]\right)$$

where we have applied Slutsky's lemma after using Assumption 12(2) to control the second term. By the result of example 3.10.6 in Van der Vaart and Wellner (1996) this means that the desired contiguity result holds, and we have that for $LR = \log\left(\prod_{t=1}^{T} \frac{f(X_t|\theta_0,Z_t)+\frac{1}{\sqrt{T}}d(X_t,Z_t)}{f(X_t|\theta_0,Z_t)}\right)$, $LR \to_d N\left(-\frac{1}{2}E_{\theta_0}\left[\frac{d(X_t,Z_t)^2}{f(X_t|\theta_0,Z_t)^2}\right], E_{\theta_0}\left[\frac{d(X_t,Z_t)^2}{f(X_t|\theta_0,Z_t)^2}\right]\right)$. By another application of the Martingale Central Limit Theorem, under Assumption 7 this implies that for any finite collection $(\lambda_1, \xi_1)...., (\lambda_k, \xi_k) \in \Lambda\times\Xi$, we have that the vector $(\sqrt{T}R(\theta_0, \lambda_1, \xi_1), ..., \sqrt{T}R(\theta_0, \lambda_k, \xi_k), \gamma_0, LR)$ converges jointly to a normal distribution under $F(x|\theta_0, z)$. By an application of Le Cam's third lemma and Assumption 12(4), this implies that under $F_T(\cdot|Z_t)$ we have finite-dimensional convergence

$$\begin{bmatrix} \sqrt{T}R(\theta_0, \lambda_j, \xi_j) \\ \sqrt{T}\gamma_0 \end{bmatrix} \Rightarrow \begin{pmatrix} \mathbb{G}_1(\lambda_j, \xi_j) + \sqrt{T_0}E_{f_0}\left[\left(\frac{g(X_t|\lambda_j,Z_t)}{f(x|\theta_0,Z_t)} - 1\right)\psi(\Phi(Z_t)'\xi_j)\right] \\ \eta_0 + \sqrt{T_0}\int\int \gamma(x, z, \theta_0)f_0(x|z)d\mu(x)dF_Z \end{pmatrix}.$$

To extend this to functional convergence, we need only prove that $\begin{bmatrix} \sqrt{T}R(\theta_0, \lambda, \xi) \\ \sqrt{T}\gamma_0 \end{bmatrix}$ is asymptotically tight under $F_T(\cdot|Z_t)$. This follows from another application of Le Cam's Third Lemma (see Theorem 3.10.7 in Van der Vaart and Wellner (1996)). Finally, note that the result of Lemma A.1 holds under $F_T(\cdot|Z_t)$ by contiguity. Combined with the result above, this yields that

$$\sqrt{T}R(\hat{\theta}, \lambda, \xi) \Rightarrow \mathbb{G}^* + \sqrt{T_0}\nu(\lambda, \xi)$$

as desired.□

Acknowledgments The author is grateful to Anna Mikusheva and Whitney Newey for their guidance and support, and to Victor Chernozhukov, Denis Chetverikov, Jerry Hausman, and the participants of the MIT Econometrics Lunch for helpful comments. NSF Graduate Research Fellowship support under grant number 1122374 is gratefully acknowledged.

References

Andrews, D., 1997, A Conditional Kolmogorov Test. Econometrica 65, 1097-1128.

Bai, J., 2003, Testing Parametric Conditional Distributions of Dynamic Models. Review of Economics and Statistics 85, 531-549.

Bierens, H., 1982, Consistent Model Specification Tests. Journal of Econometrics 20, 105-134.

Bierens H. and L. Wang, 2012, Integrated Conditional Moment Tests for Parametric Conditional Distributions. Econometric Theory 28, 328-362.

Chesher, A., 1984, Testing for Neglected Heterogeneity. Econometrica 52, 865-872.

Chesher, A., D. Dhaene, C. Gourieroux and O. Scaillet, 1999, Bartlett Identities Tests. Unpublished manuscript.

Hausman, J., 1978, Specification Testing in Econometrics. Econometrica 46, 1251–1271.

Inoue, A., 1998, A Conditional Goodness-of-Fit Test for Time Series. Unpublished manuscript.

Lehmann, E.L. and J.P. Romano, 2005, Testing Statistical Hypotheses, 3rd edition. New York, Springer.

Newey, W., 1985, Maximum Likelihood Specification Testing and Conditional Moment Tests. Econometrica 53, 1047-1070.

Nishiyama, Y., 2000, Weak Convergence of Some Classes of Martingales with Jumps. Annals of Probability 28, 685-712.

Stinchcombe, M.B. and H. White, 1998, Consistent Specification Testing with Nuisance Parameters Present Only Under The Alternative. Econometric Theory 14, 295-325.

Tsirelson, V. S., 1976, The Density of the Distribution of the Maximum of a Gaussian Process. Theory of Probability and its Applications 20, 207-216.

Van der Vaart, A.W. and J.A. Wellner, 1996, Weak Convergence and Empirical Processes. New York, Springer.

White, H., 1980, A Heteroskedasticity-Consistent Covariance Matrix Estimator and A Direct Test For Heteroskedasticity. Econometrica 48, 817-838.

White, H., 1982, Maximum Likelihood Estimation of Misspecified Models. Econometrica 50, 1-25.

Zheng, J., 2000, A Consistent Test of Conditional Parametric Distributions. Econometric

Theory 16, 667-691.

Tests	$H_Z^{(0)}$			$H_Z^{(1)}$			$H_Z^{(2)}$			$H_Z^{(3)}$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\text{MLR}_Z^{(N)}$	1.2	4.1	7.2	3.3	10.3	17.1	18.6	34.8	41.1	59.4	82.1	87.2
$\text{MLR}_Z^{(0)}$	1	4.3	7.4	6.1	14.9	25.2	24.9	43.7	50	55.8	75.5	78.8
$\text{MLR}_Z^{(1)}$	1.2	4.3	7.2	16	28.5	33.7	46.5	68.1	73.8	77.2	95.4	96.4
$\text{MLR}_Z^{(2)}$	1	4.4	7.4	9.7	17.9	26.2	41.6	58.9	63.6	73	90.7	93.2
$\text{MLR}_Z^{(3)}$	0.5	5.1	9	2	6	11.5	11.4	19.3	27.2	73.4	90.8	96.1
CK	1.2	5.7	10.6	3.7	11.2	20.0	6.4	19.8	28.5	99.8	100	100
Z(0.5)	1.3	2.2	4.6	12.4	15.1	17.2	31.6	40.2	46.1	95.1	95.3	95.3
Z(1.0)	1.7	2.5	4.0	12.1	14.2	16.7	43.7	55	59.5	96.3	96.6	96.7
Z(2.0)	2.8	3	4.7	18.7	22.8	25.6	55.5	66.6	74	89.8	93.1	94
BW c=5	1.5	4.8	10.5	2.1	8.2	15.1	6.6	17.5	26.6	100	100	100
BW Max	1.6	5.2	9.7	1.8	6.9	13.7	5.4	16	22.7	100	100	100

Table 1: Simulation results in linear regression model. The first three columns report size in percent of nominal 1%, 5%, and 10% tests, while remaining columns report power against alternatives $H_Z^{(1)} - H_Z^{(3)}$ specified in (4). $\text{MLR}_Z^{(i)}$ denotes the mean likelihood ratio tests, CK is Conditional Kolmogorov Test of Andrews (1997), Z(b) is Zheng (2000)'s test with bandwidth choice b, BW c=5 is Bierens and Wang (2012) test with integration domain equal to a cube around zero with side length 10, and BW Max is Bierens and Wang test maximized over the integration domain. Results for the non-MLR tests are taken from Bierens and Wang (2012).

Tests	$H_P^{(0)}$			$H_P^{(1)}$			$H_P^{(2)}$			$H_P^{(3)}$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\text{MLR}_P^{(N)}$	0.6	3.1	8.2	54.1	63.9	65.7	77.9	94	96.4	45.8	65.7	67.6
$\text{MLR}_P^{(0)}$	0.7	3.4	6.9	64.5	78.8	79.4	91.6	98.5	98.5	70.9	76	76.7
$\text{MLR}_P^{(1)}$	0.9	3.3	7.4	97.7	99.2	99.2	97.6	100	100	93.4	96.9	97.1
$\text{MLR}_P^{(2)}$	0.7	4.8	8.9	95.6	98.4	98.4	98.6	99.6	99.6	96.4	99.6	100
$\text{MLR}_P^{(3)}$	0.8	6.3	11.8	95.1	97.4	97.4	97.9	99.3	99.3	98.5	99.7	99.7
CK	0.5	5.6	11.4	98.3	99.3	99.6	53.6	74.3	83.7	21.1	47.5	61.8
BW c=5	1.7	6.0	10.7	75.3	89.8	94.8	69.2	88.9	94.3	47.0	76.9	88.4
BW Max	1.0	5.2	9.6	57.8	77.7	85.4	36.5	55.9	67.2	29.1	55.6	68.7

Table 2: Simulation results in Poisson regression model. First three columns report size in percent of nominal 1%, 5%, and 10% tests, while remaining columns report power against alternatives $H_Z^{(1)} - H_Z^{(3)}$ specified in (5). $\text{MLR}_P^{(i)}$ denotes the mean likelihood ratio tests, CK is Conditional Kolmogorov Test of Andrews (1997), BW c=5 is Bierens and Wang (2012) test with integration domain equal to a cube around zero with side length 10, and BW Max is Bierens and Wang test maximized over the integration domain. Results for the non-MLR tests are taken from Bierens and Wang (2012).

Tests	$H_{AR}^{(0)}$			$H_{AR}^{(1)}$			$H_{AR}^{(2)}$			$H_{AR}^{(3)}$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\text{MLR}_{AR}^{(N)}$	0.9	4.4	7.8	7	17.9	27.1	26.5	43.2	53.8	94.1	98.1	99.3
$\text{MLR}_{AR}^{(0)}$	1.1	4.3	9.1	7	16.9	25.3	9.5	19.7	25.6	99.1	99.9	99.9
$\text{MLR}_{AR}^{(1)}$	0.8	4	7.1	21.1	37	41.9	53.4	72.8	75.5	95.8	96	96.1
$\text{MLR}_{AR}^{(2)}$	1.1	4.1	7.8	14.7	26.5	34.1	51.5	63.3	66.4	99.2	99.4	99.8
$\text{MLR}_{AR}^{(3)}$	1.2	5.7	11.4	14.2	29.1	32.7	46.8	61.6	63.8	99.9	99.9	99.9

Table 3: Simulation results in autoregressive model. First three columns report size in percent of nominal 1%, 5%, and 10% tests, while remaining columns report power against alternatives $H_Z^{(1)} - H_Z^{(3)}$ specified in (6). $\text{MLR}_{AR}^{(i)}$ denotes the mean likelihood ratio tests.