Super-Consistent, Non-Degenerate Tests of Functional Form: Optimal CM Tests when Nuisance Parameters are Present under the Null

[Under Revision]

Jonathan B. Hill
Dept. of Economics
Florida International University
October 5, 2004

Abstract

This paper develops parametric conditional moment tests of functional form that are either (i) asymptotically non-degenerate for any non-zero nuisance vector $\gamma$; or (ii) consistent for any non-zero nuisance vector $\gamma$; or (iii) both. In cases (i) and (iii), our moment conditions will help improve upon certain "optimal" tests developed by Andrews and Ploberger (1994); in the latter cases (ii) and (iii), the tests provide a simple means to bypass the power leverage implied by the Integrated Conditional Moment (ICM) tests of Bierens (1982) and Bierens and Ploberger (1997).

Moreover, we prove that test weights $F(\gamma \Psi(x))$ may incorporate functional mappings $\Psi(x)$ with far more structure than a property of one-to-one allows, suggested by Bierens (1990). An appropriate choice of the argument $\Psi(x)$ alone is enough to lead to a consistent and non-degenerate test for every $\gamma \neq 0$, where computation requires projecting the test statistic onto an integer set rather than a compact subset of the reals, as is typically required.

Furthermore, consistent and/or non-degenerate tests for every $\gamma \neq 0$ are developed for general $\Psi(x)$ functionals. The implied alternative models are hybrid forms of smooth transition type structures: an amalgamation of functional forms popularized in the neural networks, regime switching and structural change literatures leads to "super"-consistent and "super" optimal, non-degenerate tests of functional form.

*Dept. of Economics, Florida International University, Miami, Fl; jonathan.hill@fiu.edu; http://www.fiu.edu/~hilljona
JEL classification: C12, C45, C52
Key words: consistent conditional moment test; nuisance parameter; Bierens test.
1. **Introduction** In this paper we develop conditional moment tests of functional form for time series models that are consistent against any deviation from the null specification, and variously solve fundamental shortcomings of the parametric tests developed by Bierens (1990), Bierens and Andrews (1994), and Bierens and Ploberger (1997). The asymptotic distributions of the test statistics depend on nuisance parameters, say \( \gamma \), which are not defined under the null. Examples include tests of continuous structural breaks, parameter instability and threshold effects. In particular, the conditional moments developed here variously lead to Lagrange Multiplier tests that are consistent (asymptotic power of one) against any deviation from the null for every non-zero value of the nuisance vector; non-degenerate for every non-zero nuisance vector; or both.

Conditional moment (CM) tests based on a finite number of \( L_2 \)-orthogonality conditions, cf. Newey (1985) and Tauchen (1985), are known not to be consistent against any alternative. Apparently, however, the only consistent parametric\(^1\) tests are those of Bierens (1982, 1990) and Bierens and Ploberger (1997). As such, the "Bierens test" has become the parametric standard by which consistent functional specification tests are measured. Consider a functional specification for some stochastic process \( \{y_t\} \),

\[
y_t = f(x_t, \theta) + \epsilon_t
\]

(see Section 2 for details). The null hypothesis of interest states the function \( f(x_t, \theta) \) is for some \( \theta_0 \) the expectations \( E[y_t|x_t] \) with probability one. If the null functional form \( f(x_t, \theta) \) is mis-specified for any \( \theta \), then under strikingly general conditions the following orthogonality condition holds,

\[
E[(y_t - f(x, \theta))e^{\gamma \Psi(x_t)}] \neq 0,
\]

for any bounded one-to-one function \( \Psi(x_t) : \mathbb{R}^k \to \mathbb{R}^k \), except for a countable set \( S \) of nuisance vectors \( \gamma \in \mathbb{R}^k \): see Bierens (1990: Lemma 1), and see de Jong (1996) for the case of stationary, infinite dimensional dependent processes \( \{y_t, x_t\} \).

Stinchcombe and White (1998) characterize the class of "generically comprehensive" functions \( F(\gamma \Psi(x_t)) \) which have the above property (2); essentially any real analytic function which converges on a compact subset of \( \mathbb{R} \) on which \( \gamma \Psi(x_t) \) takes it values, including the exponential, logistic and \( \sin + \cos \). See also Bierens and Ploberger (1997).

The set \( S \) in general depends on the data generating process \( \{y_t, x_t\} \) and therefore is typically unknown in practice. It is nonetheless interesting to point out that nowhere in the literature does there appear to be a discussion about the contents of \( S \) other than non-dense countability, and whether test performance can be enhanced based on "manipulating" or "forcing" the contents of \( S \).

\(^1\)For consistent nonparametric tests, see, e.g., Lee (1988), Yatchew (1992), Wooldridge (1992), Zheng (1996), and Hong and White (1995). Non-parametric methods are suitable for testing whether a particular functional specification is correct (with probability one), but cannot provide a better parametric specification if the null specification is false, and therefore will not be considered here.
Similarly, an LM statistic based on a sample version of (2) requires a non-degenerate asymptotic variance, say \( V(\gamma) > 0 \). Bierens (1990: Lemma 2) proves the set \( S^* \) of \( \gamma \) for which the variance is zero is countable. In practice, of course, we cannot know if a particularly chosen nuisance vector \( \gamma \) will render inconsistent and/or degenerate the asymptotic properties of the test statistic. Indeed, for general sample moment conditions of the form

\[
\hat{s}(\gamma) = \frac{1}{n} \sum_{t=1}^{n} (y_t - f(x_t, \hat{\theta})) h(x_t, \delta) F(\gamma' \Psi(x_t)),
\]

where \( h(x_t, \delta) \) may be vector-valued with some vector \( \delta \) of parameters, Bierens (1990) claims we will have to assume the resulting asymptotic covariance matrix is \( \sqrt{n} \)-nonsingular for all \( \gamma \neq 0 \) (see also de Jong, 1996), and claims (p. 1446) the choice of weighting vector \( h(x_t, \delta) \) is "irrelevant for consistency".

In this paper, we characterize a class of weights \( h(x_t, \delta) \) which variously solve each shortcoming of the Bierens test, and thereby effectively control the contents of \( S \) and/or \( S^* \): the weights imply either (i) consistency for all nuisance vectors \( \gamma \neq 0 \) (i.e. \( S = \{0\} \)); or (ii) a positive definite asymptotic covariance matrix \( V(\gamma) \) for every nuisance vector \( \gamma \neq 0 \), and in some cases \( \gamma = 0 \) (i.e. \( S^* = \{0\} \) or \( \emptyset \)); or (iii) both consistency and non-degeneracy for every \( \gamma \neq 0 \). We refer to the property of test consistency for any nuisance vector \( \gamma \neq 0 \) as "super-consistency". In some cases, super-consistency can even hold at \( \gamma = 0 \).

Thus, the choice of weighting function \( h(x_t, \delta) \) is neither irrelevant for matters of consistency if coverage over the entire nuisance parameter space is desired, nor necessarily prompts the assumption of test statistic non-degeneracy. Indeed, if we simply set \( h(x_t, \delta) = \partial f(x_t, \theta) / \partial \theta \), then a consistent test statistic that is non-degenerate for any \( \gamma \neq 0 \) is available, where the weight \( \partial f(x_t, \theta) / \partial \theta F(\gamma' \Psi(x_t)) \) directs power toward "smooth transition" type models, a la Teräsvirta (1994).

Furthermore, we show (iv) that the argument \( \Psi(x_t) \) need not be one-to-one and may be parametric. For example, a test of linear autoregression against an exponential smooth transition autoregression (ESTAR) will be both non-degenerate and consistent (but not necessarily super-consistent) despite the implied argument \( \Psi(x_t) \) being parametric and non-one-to-one. Indeed, for a test of linearity simply choosing the functional \( h(x_t, \delta) = x_t \) and argument \( \Psi(x_t(\delta)) = (\exp(\delta_1 x_{t,1}), \ldots, \exp(\delta_k x_{t,k}))' \) for any \( \delta_i \neq 0, i = 1 \ldots k \), leads to infinitely many super-consistent and non-degenerate test statistics with respect to nonzero \( \gamma \). The test statistic in this case depends on a nuisance integer set \( m \), thus computation requires projecting the statistic onto a known, countable set \( \{1, 2, \ldots\} \) rather than an arbitrary compact sub-set of the reals, as is typically required (e.g. Bierens, 1982, 1990; Bierens and Ploberger, 1997).

As such, Integrated Conditional Moment (ICM) tests, cf. Bierens (1982) and Bierens and Ploberger (1997), may not have a comparative power advantage over the tests developed under (i) and (iii), and (iv). ICM tests are typically

---

2Whether the tests developed here provide a power advantage over ICM tests is a question of whether empirical power converges to one faster then for ICM tests. We leave this question for future research. Our point here is that integrating the moment condition itself in order to fish out a non-zero moment is superfluous if all moments are non-zero under the alternative.
constructed from sample moments based on $\int (E[\epsilon_t F(\gamma' \Psi(x_t))]^2 \mu(\gamma))$ for some absolutely continuous weighting function $\mu(\gamma)$. Because $E[\epsilon_t F(\gamma' \Psi(x_t))] \neq 0$ for some $\gamma$ under $H_1$, the integrated squared moment is guaranteed to be positive under the alternative. In the present paper we prove under (i) and (iii) the moment condition $E[\epsilon_t h(x_t, \delta) F(\gamma' \Psi(x_t))] \neq 0$ for any $\delta$ and non-zero $\gamma$, hence the choice of $\gamma$ is irrelevant, and we can safely by-pass the computational task of numerical integration in practice and the essentially arbitrary choice of measure $\mu(\gamma)$.

However, integration of the LM statistic itself with respect to $\gamma$, $\int LM(\gamma) d\mu(\gamma)$, is always the limit of an optimal test that effectively directs power toward small deviations from the null: see Andrews and Ploberger (1997). Of course, if $\det V(\gamma) = 0$ at some point $\gamma$, then even the average statistic $\int LM(\gamma) d\mu(\gamma)$ will be asymptotically degenerate, unless $\mu(\gamma)$ places zero weight on those points, which are in general unknown. Indeed, each of the "optimal" LM-statistics developed in Andrews and Ploberger (1994), including the limiting sup-statistic, require non-degeneracy over all points $\gamma$ where $\mu(\gamma)$ places non-zero measure, a requirement that is not guaranteed in the Bierens (1990) framework, cf. (2). The moment conditions developed under (ii)-(iv), however, are always non-degenerate for $\gamma \neq 0$, thus LM tests based on those conditions will even improve upon the "optimality" of the class of optimal LM statistics a la Andrews and Ploberger (1994).

Tests based on (3) are simply orthogonality tests of the omitted structure $h(x_t, \delta) F(\gamma' \Psi(x_t))$. Under the alternative the optimally selected models are similar in spirit to smooth transition autoregressions, cf. Teräsvirta (1994), with multiple regimes and variously hybrid forms of neural networks, cf. Kuan and White (1994) and Bierens (1990), regressor cross-products, cf. White (1980), and stochastic coefficients. Thus, an exact combination of the most promising functional forms used in the econometrics literatures to date provides fail-safe Bierens-type functional mis-specification tests.

In order to reduce notation, we restrict attention to ergodic, stationary time series processes $\{y_t, x_t\}$ with finite dimensional regressors, $x_t$. However, it seems straightforward to extend the results here to the infinite dimensional $x_t$ (e.g. ARIMA) case in a manner similar to de Jong (1996). In practice, $x_t$ may contain a constant, lags of $y_t$ and exogenous information; $y_t$ may be vector-valued (e.g. VAR); and $\{y_t\}$ may denote a squared innovations process in the manner of stationary ARCH models. For simplicity and continuity, examples provided below are for autoregressive process and often for linear null specifications. We also provide numerous demonstrations of how the theory developed here can be extended to tests of smooth transition nonlinearity.

In Section 2, we construct a basic vector moment condition test, and develop a class of optimal weights in Sections 3 and 4. Section 5 contains examples of models implied by logistic weighted tests. In Section 6 we perform a simulation study comparing the performance of the tests developed here against the extant consistent parametric tests of Bierens (1990) and Lee et al (1993). Concluding remarks are in Section 7, assumptions can be found in Appendix 1, Appendix 2 contains tables, and all proofs are left for Appendix 3.
Throughout we maintain the following notation conventions. \( \rightarrow \) denotes convergence in probability; \( \implies \) denotes weak convergence with respect to finite dimensional distributions. \( | \cdot | \) denotes the Euclidean norm for real-valued vectors, and the matrix norm for real-valued square matrices: \( |x| = [\text{Tr}(x'x)]^{1/2} \).

Denote by \( \perp \) orthogonality in \( L_2(\Omega, \mathcal{F}_t, P) \), where \( \mathcal{F}_t \) denotes an increasing \( \sigma \)-algebra induced by \( \{y_t, x_t\}, \{y_{t-1}, x_{t-1}\}, \ldots \), and let \( (x, y) \) denote the Euclidean inner product. Let \( \mathbb{F}(z_1 : 1 \leq i \leq n) \) denote the closed linear span of the sequence \( \{z_i\}_{i=1}^n, n \geq 1 \). For arbitrary \( k \)-vectors \( a \) and \( x \), vector powers \( x^n \) are understood to represent \( (x_1, \ldots, x_k)' \). Denote by \( \mathbb{Z} \) the set of integers: \( \mathbb{Z} = \{\ldots -2, -1, 0, 1, 2, \ldots\} \). We use \( m_i \in \mathbb{Z}_+ = \{0, 1, 2, \ldots\} \), to denote arbitrary integers with arbitrary indices \( i = 1, 2, \ldots \), whose values are understood in context.

We write \( I_m \) to denote an \( m \)-dimensional identity matrix.

2. **Vector-Valued Conditional Moments**

Let \( \{y_t, x_t\} \) be a jointly distributed strictly stationary, ergodic time series for each \( t \), where \( E|y_t| < \infty \) and \( x_t \) is an \( \mathcal{F}_{t-1} \)-measurable \( k \)-vector. The Radon-Nikodym theorem guarantees

\[
E[y_t|x_t] = g(x_t), \text{ a.s.} \tag{4}
\]

for some \( \mathcal{F}_{t-1} \)-measurable function \( g(x_t) : \mathbb{R}^k \rightarrow \mathbb{R} \). We assume the function \( g(x_t) \) is a member of a parametric family of known real functions \( f(x_t, \theta) : \mathbb{R}^k \times \Theta \rightarrow \mathbb{R} \), where \( \Theta \) denotes a compact, convex subset of \( \mathbb{R}^k \). Define \( \epsilon_t \equiv y_t - f(x_t, \theta) \), and write tautologically

\[
y_t = f(x_t, \theta) + \epsilon_t. \tag{5}
\]

The fundamental hypotheses are

\[
H_0 : P(E[y_t - f(x_t, \theta_0)|x_t] = 0) = 1, \tag{6}
\]

for some \( \theta_0 \in \Theta \), and

\[
H_1 : \sup_{\theta \in \Theta} P(E[y_t - f(x_t, \theta)|x_t] = 0) < 1. \tag{7}
\]

Under \( H_0 \) there exists some set \( \theta_0 \) such that \( f(x_t, \theta_0) \) is almost surely correctly specified, and \( \epsilon_t \) forms a martingale difference sequence. The alternative \( H_1 \) is simply that the null model is mis-specified, hence \( H_1 \) embraces any deviation from the null.

2.1 **Vector Moment Conditions**

Define the closed, bounded, compact parameter subspaces \( \Gamma \subseteq \mathbb{R}^k \) and \( \Delta \subseteq \mathbb{R}^{m_0 \times m_1} \) for arbitrary \( m_0, m_1 \in \mathbb{Z}_+ \). Denote by \( h(x_t, \delta) \) any uniformly bounded, continuous mapping from \( \mathbb{R}^k \times \Delta \) to \( \mathbb{R}^{m_2} \), \( m_2 \geq 1 \), measurable with respect to \( \mathcal{F}_{t-1} \), such that

\[
0 < \inf_{\delta \in \Delta} |h_i(x_t, \delta)| \leq \sup_{\delta \in \Delta} |h_i(x_t, \delta)| < \infty \text{ with probability one for each } i = 1 \ldots m_2. \tag{8}
\]

In general, \( h(x_t, \delta) \) may be non-zero constant valued. For example, if \( 0 \in \Delta \), then \( h(x_t, \delta) \) cannot be \( \delta' x_t \) because \( \delta' x_t = 0 \) with probability one when \( \delta = 0 \); however \( h(x_t, \delta) = \exp(-\delta' x_t^2) \) is plausible for bounded

---
real vectors $\delta \geq 0$. We assume $h(x_t, \delta)$ is uniformly twice continuously differentiable in $\delta$ for estimation asymptotics, however the main moment condition results of this paper do not rely on this assumption.

The following lemma provides a direct extension of Theorem 1 of Bierens and Ploberger (1997) to generalized vector moment conditions, and helps us to analyze the required restrictions on the weight argument $\Psi(x)$. For consistency for some nuisance vector $\gamma$, we will only require test weights $F(u) = F(\gamma' \Psi(x_t))$ to be infinitely many times continuously differentiable at $u = 0$ such that the set

$$\{s \in \mathbb{Z} : (d/du)^s F(u)|_{u=0} = 0 \text{ with positive measure} \}$$

is finite, thus $F(u)$ cannot be polynomial: see Theorems 2.3 and 3.10 of Stinchcombe and White (1998). However, we maintain the following compound assumption in order to sharpen the main results.

Assumption A. The weight $F(u)$ satisfies (8), $(d/du)^s F(u) \neq 0$ with probability one, and $F(u)$ is a convergent power series on an open interval $R_0$ of the real line with closure containing 0:

$$\forall u \in R_0 \subset \mathbb{R} : F(u) = \sum_{s=0}^{\infty} [F^s(0)/s!] u^s$$

where $F^s(0) = (d/du)^s F(u)|_{u=0}$. For example, the exponential $F(u) = \exp(u)$ or logistic $F(u) = (1 + \exp(c-u))^{-1}$ hold, provided $c \neq 0$ for the latter. Thus, $F(u)$ is real analytic.

Notice that for any functions $F(u)$ and $G(\tilde{u})$ that satisfy Assumption A, in general infinitely many $s^{th}$-derivatives $(\partial/\partial u)^s F(u)$ and products $F(u)G(\tilde{u})$ also satisfy Assumption A. For example, if $G(\gamma' x) = \exp(\gamma' x)$ and $F(\gamma' x) = \exp(\gamma' x)$, then $F(\gamma' x)G(x, \tilde{\gamma})$ satisfies Assumption A if and only if $\tilde{\gamma} \neq -\gamma$. Consult Appendix 1 for all assumptions, and Appendix 3 for proofs.

For compactness, define $\hat{F}_{t,i}(\gamma, \delta) \equiv h_i(x, \delta)F(\gamma' x)$.

Lemma 1. Let $e$ be a random variable satisfying $E|e| < \infty$, and let $x$ be an $\mathcal{F}$-measurable bounded vector in $\mathbb{R}^k$, $0 < k < \infty$, such that $P[E(e|x) = 0] < 1$. Assume $\Gamma$ is a bounded subset of $\mathbb{R}^k$, and let Assumption A hold for $F(\gamma' x)$ and $\hat{F}_{t}(\gamma, \delta)$. Then for each $\delta \in \Delta$ the sets

$$S_i = \{\gamma \in \Gamma : E[e\hat{F}_{t,i}(\gamma, \delta)] = 0 \text{ and } P(\gamma' x \in R_0) = 1\}, \ i = 1...k,$$

have Lebesgue measure zero, and are nowhere dense in $\mathbb{R}^k$.

Remark 1: Conditioning on $x$ in $E(e|x)$ is equivalent to conditioning on any bounded, measurable, one-to-one function of $x$, say $\Psi(x) : \mathbb{R}^k \rightarrow \mathbb{R}^k$, since any such functional induces the same $\sigma$-field as $x$: see Billingsley (1995: Theorem 5.1). Thus, it is not strictly necessary to assume $x$ is bounded as long as a bounded functional $\Psi(x)$ is used. See also Bierens (1990).

Remark 2: The fact that the conditional moment $E[e\hat{F}_{t,i}(\gamma, \delta)] \neq 0$ under $H_1$ holds for any $\delta \in \Delta$ supports Bierens’ (1990) claim that such weights $h(x, \delta)$
are irrelevant for consistency. We can, nonetheless, select weights \( h(x, \delta) \) such that \( S_i = \{0\} \); see Section 3.

**Remark 3:** The requirement that \( \hat{F}_{i, i}(\gamma, \delta) = h_i(x, \delta) F(\gamma' x) \) satisfy Assumption A can be easily defended: we must restrict \( h_i(x, \delta) \) not effectively to "cancel out" \( F(\gamma' x) \), such as \( h_i(x, \delta) = d_i \times F(\gamma' x)^{-1} \) for any scalar constant \( d_i \).

From Lemma 1, Assumptions 1–4 and standard asymptotic theory, the following distribution limit holds.

**Theorem 2** Let Assumptions 1–4 and Assumption A hold. Then, (i) under \( H_0 \)

\[
\sqrt{n} \hat{s}(\gamma, \delta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (y_i - f(x_i, \hat{\theta})) \hat{F}_{i, i}(\gamma, \delta) \Rightarrow N(0, V(\gamma, \delta)),
\]

point-wise in \( \gamma \in \Gamma \) and \( \delta \in \Delta \), with asymptotic covariance matrix

\[
V(\gamma, \delta) = E \left[ z^2(\gamma, \delta) z(\gamma, \delta)' \right],
\]

where

\[
\begin{align*}
\hat{F}_{i, i}(\gamma, \delta) - b(\gamma, \delta) A^{-1} \partial \partial \theta f(x_i, \theta_0) \\
b(\gamma, \delta) = E \left[ \hat{F}_{i, i}(\gamma, \delta) \partial \partial \theta f(x_i, \theta_0) \right] \\
A = E \left[ \partial \partial \theta f(x_i, \theta_0) \partial \partial \theta f(x_i, \theta_0) \right].
\end{align*}
\]

Moreover, (ii) under \( H_1 \) there exists a subset \( S \) of \( \mathbb{R}^k \) with Lebesgue measure zero such that for any \( \delta \in \Delta \) and each \( \gamma \in \Gamma / S \)

\[
\hat{s}(\gamma, \delta) = \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i, \hat{\theta})) \hat{F}_{i, i}(\gamma, \delta) \rightarrow \eta_i(\gamma, \delta)
\]

with probability one for some vector-functional \( \eta(\gamma, \delta) \) with scalar components \( \eta_i(\gamma, \delta) \neq 0 \) for each \( i = 1...k \).

**3 Conditioning Set \( \Psi(x) \)** Lemma 1 can be used to demonstrate that the fact that we only require \( h(x, \delta) \) to be bounded and non-zero with probability one, and \( h(x, \delta) F(\gamma' \Psi(x)) \) to satisfy Assumption A, implies the compound weight \( h(x, \delta) F(\gamma' \Psi(x)) \) can have far more structure than \( F(\gamma' \Psi(x)) \) alone. Indeed, \( \Psi(x) \) need not be one-to-one, and may be parametric. Moreover, if we choose a parametric mapping \( \Psi(x, \delta) = (\exp(\delta_1 x_1), ..., \epsilon(\delta_k x_k))' \) for any bounded \( k \)-vector \( \delta \in \Delta, \delta_i \neq 0, i = 1...k \), then a moment condition that is super-consistent relative to any \( \gamma \neq 0 \) in a compact subset of the reals is available.

**3.1 Conditioning Set \( \Psi(x) \): Properties**

Subject to Lemma 1, evidently the only restriction we need to impose on any scalar weight \( F(\gamma' \Psi(x)) \) that satisfies Assumption A is that it decompose into a product \( h(x, \delta) F(\gamma' \Psi(x)) \) for some bounded, one-to-one mapping \( \Psi(x) : \mathbb{R}^k \rightarrow \mathbb{R}^k \), some uniformly bounded scalar \( h(x, \delta) \neq 0 \) with probability one, and some function \( F(u) \) that satisfies Assumption A. We summarize this result in the following assumption and lemma.
Example 1: ESTAR

Let traders of financial assets respond asymmetrically to the timing and content of news. See, also, Pagan and Schwert (1990) and Ding (1976), that traders of financial assets respond asymmetrically to the timing and content of news. See, also, Pagan and Schwert (1990) and Ding (1976).

\[ F(\gamma^\prime \Psi(x)) \]

Assumption A' The weight \( \tilde{F}(\gamma^\prime \Psi(x)) \) satisfies \( \tilde{F}(\gamma^\prime \Psi(x)) = h(x, \delta)F(\gamma^\prime \Psi(x)) \), where \( F(\gamma^\prime \Psi(x)) \) satisfies Assumption A, \( \Psi(x) \) is a measurable, bounded, one-to-one function of \( x \), and the scalar function \( h(x, \delta) : \mathbb{R}^k \times \Delta \rightarrow \mathbb{R} \) is bounded and non-zero with probability one.

Lemma 3 Let \( e \) be a random variable satisfying \( E|e| < \infty \), and let \( x \) be an \( \mathbb{R}^k \)-measurable bounded vector in \( \mathbb{R}^k \), \( 0 < k < \infty \), such that \( P(E|e| = 0) < 1 \). Assume \( \Gamma \) is a bounded subset of \( \mathbb{R}^k \), and let Assumption A' hold for some \( \tilde{F}(\gamma^\prime \Psi(x)) \). Then the set

\[ S = \{ \tilde{\gamma} \in \Gamma : E[\epsilon \tilde{F}(\gamma^\prime \Psi(x))] = 0 \text{ and } P(\gamma^\prime \Psi(x) \in R_0) = 1 \}, \]

has Lebesgue measure zero, and is nowhere dense in \( \mathbb{R}^k \).

Example 1: ESTAR A score test of linear autoregression against an exponential smooth transition autoregressions (ESTAR) would use the weight \( \tilde{F}(\gamma^\prime \Psi(x, c)) = x \exp(-\sum_{i=1}^k \tilde{\gamma}_i |x_i - c_i|^2) \), for some set of constant "thresholds" \( c_i \) and transition scales \( \tilde{\gamma}_i \geq 0 \); see, e.g., Teräsvirta (1994), van Dijk et al (2000), and Hill (2004).

If the thresholds are assumed to satisfy \( c_i = 0 \) for each \( i = 1...k \), then

\[ \tilde{F}(\gamma^\prime \Psi(x, c)) = x \exp(-\sum_{i=1}^k \tilde{\gamma}_i |x_i|^2) = h(x, \delta)F(\gamma^\prime \Psi(x)) \]

where \( h(x, \delta) = x \) and \( F(\gamma^\prime \Psi(x)) = \exp(-\sum_{i=1}^k \tilde{\gamma}_i |x_i|^2) \), hence \( \Psi(x) = -x^2 \), a non-one-to-one function of \( x \), and either Lemma 1 or 3 fails to hold.

An empirical example of such a problematic moment condition lies with testing smooth transition ARCH models (STARCH); see, e.g., Gonzalez-Rivera (1998). In the STARCH literature often the threshold mechanism is assumed to be grounded on the sign of the underlying innovations\(^3\), \( \epsilon_t \), therefore the thresholds \( c_i \) are assumed to be zero.

However, we can successfully decompose \( \tilde{F}(\gamma^\prime \Psi(x, c)) \) for most values of \( \tilde{\gamma} \) and \( c \). Assume \( x \) contains a constant term, and write \( \tilde{F}(\gamma^\prime \Psi(x, c)) = x \exp(-\sum_{i=1}^k \tilde{\gamma}_i |x_i - c_i|^2), \tilde{\gamma}_i \geq 0 \), as

\[ \tilde{F}(\gamma^\prime \Psi(x, c)) = x \exp\left( -\sum_{i=1}^k \tilde{\gamma}_i |x_i - c_i|^2 \right) \]  

\[ = x \exp\left( -\sum_{i=1}^k \tilde{\gamma}_i x_i^2 - 2x_i c_i + c_i^2 \right) \]

\[ = \omega \times x \exp\left( -\sum_{i=1}^k \delta_i x_i^2 - \sum_{i=1}^k \gamma_i x_i \right) \]

\[ = \omega \times x \exp(-\delta^\prime x^2) \exp(-\gamma^\prime x) \]

\[ = \omega \times h(x, \delta) \times F(\gamma^\prime \Psi(x)) \]

\(^3\)There exists ample empirical evidence, noted early by Mandelbrot (1963) and Black (1976), that traders of financial assets respond asymmetrically to the timing and content of news. See, also, Pagan and Schwert (1990) and Ding et al (1993).
where we define

$$\omega = \exp \left( \sum_{i=1}^{k} \hat{\gamma}_i c_i \right);$$  \hspace{1cm} (17)

$$\hat{\Psi}(x, c) = -[(x_1 - c_1)^2, ..., (x_k - c_k)^2]^\prime;$$

$$h(x, \delta) = x \exp(\delta' x^2); \quad F'(\gamma \hat{\Psi}(x)) = \exp(\gamma' x);$$

$$\delta = \hat{\gamma} \geq 0; \quad \gamma_i = -2\hat{\gamma}_i c_i, i \geq 1.$$

We factor \(\omega = \exp(\sum_{i=1}^{k} \hat{\gamma}_i c_i)\) out of the compound weight because it is constant-valued and therefore will add nothing to a test of linear autoregression against an ESTAR alternative. We need only use \(h(x, \delta)F'(\gamma \hat{\Psi}(x))\), with \(h(x, \delta)\) and \(F'(\gamma \hat{\Psi}(x))\) defined above. For bounded, non-zero \(h(x, \delta)\), Lemma 3 states \(E[h(x, \delta)F'(\gamma \hat{\Psi}(x))] \neq 0\) under the alternative, except for countably many \(\gamma\). Thus, for given \(\delta = \hat{\gamma}\) the stated moment condition holds except for countably many \(\gamma = [-2\hat{\gamma}_i c_i]_{i=1}^{k}\), hence for except for countably many \(c\). Therefore, in general we cannot simply fix \(c = 0\) beforehand, as is frequently the case in the ST-ARCH literature: we must allow at least one \(\hat{\gamma}_i > 0\) and \(c_i > 0\).

**Example 2: Multiplicative Thresholds**  In the STAR literature, rather than employing additive weight arguments (e.g. \(\exp(-\sum_{i=1}^{k} \hat{\gamma}_i [x_i - c_i]^2]\)), it is increasingly popular to use multiplicative forms, for example \(\exp(-\prod_{i=1}^{k} \hat{\gamma}_i [x_i - c_i]^2]\): see, e.g., Lin and Teräsvirta (1994) and He et al (2004). For any value of \(c_i, i = 1...k > 1\), however, the argument \(-\prod_{i=1}^{k} \hat{\gamma}_i [x_i - c_i]^2\) is not a one-to-one function of the threshold information \(x\) as long as more than one \(\hat{\gamma}_i > 0\), and cannot be decomposed in a manner outlined in Lemma 3. Thus, it is not likely that tests constructed to detect STAR forms of non-linearity with such functional arguments will be consistent, including the tests of continuous structural change and time-varying STAR nonlinearity developed in the above citations.

### 3.2 Conditioning Set \(\Psi(x)\): Super-Consistency on \(\mathbb{Z}^k\)

For any bounded, \(\delta \in \Delta \subseteq \mathbb{R}^k, \delta_i \neq 0, i = 1...k\), define \(x_i(\delta) \equiv (\delta_1 x_{i,1}, ..., \delta_k x_{i,k})',\) and consider the mapping \(\Psi(x_i(\delta)) = (\exp(x_{i,1}(\delta)), ..., \exp(x_{i,k}(\delta)))'.\) Recall \(F^s(u) \equiv (\partial / \partial u)^s F(u)\).

**Theorem 4** Let \(\epsilon_t\) be a random variable satisfying \(E[\epsilon_t] < \infty\), and let \(x_t\) be an \(\hat{\gamma}_{t-1}\)-measurable bounded vector in \(\mathbb{R}^k\), \(0 < k < \infty\), such that \(P[E(\epsilon_t|x_t) = 0] < 1\). Let Assumption A' hold for \(F(u)\). Then for every point \((\gamma, \delta) \in \Gamma \times \Delta, \delta_i \neq 0, i = 1...k,\) and any \(k\)-vector of integers \(r \in \mathbb{Z}^k\),

$$E \left[ \epsilon_t \exp\{m' x_t(\delta) \} \prod_{i=1}^{k} F^{m_i - r_i}(u) | u = \gamma' \Psi(x_t(\delta)) \right] \neq 0,$$

holds for infinitely many \(k\)-vectors \(m \geq r\). In particular, if \(F(u) = \exp\{u\} \) and \(r = 0\), then there exist infinitely many \(m \in \mathbb{Z}^k\) such that

$$E \left[ \epsilon_t \exp\{m' x_t(\delta) + \gamma' \Psi(x_t(\delta)) \} \right] \neq 0.$$

(19)
Remark 1: Theorem 4 holds under the general construction $\Psi(x_t(\delta)) = (\exp(\Psi(x_{t,1}(\delta))), ..., \exp(\Psi(x_{t,k}(\delta))))'$ where $\Psi(z)$ satisfies Assumption A'. In this case, (19) becomes $E[\epsilon_t \exp\{m' \Psi(x_t(\delta)) + \gamma' \Psi(x_t(\delta))\}] \neq 0$ under $H_1$ for infinitely many $m \in \mathbb{Z}_k^+$. 

Remark 2: Moment conditions (18) and (19) imply for infinitely many integers $m$ super-consistent tests are available with respect to $\gamma$ and $\delta$ in arbitrary compact subsets of $\mathbb{R}^k$, with the restriction $\delta_i \neq 0$, $i = 1...k$.

Remark 3: For brevity, consider (19) and denote by $LM(m, \gamma, \delta)$ the test statistic associated with a sample moment estimator. Because we are now only concerned about a countable set of integers, for practical purposes "removing" the nuisance integer is straightforward. For example, for any chosen $\delta_i \neq 0$, $i = 1...k$, and any $\gamma$, we need only construct an average statistic $\lim_{M \to \infty} M^{-k} \sum_m LM(m, \gamma, \delta)$, where $\sum_m$ denotes $\sum_{i=1}^{k} \sum_{m_i=0}^{M-1}$; or we can employ an average moment condition with uniform measure as a discreet version of the ICM tests of Bierens and Ploberger (1997),

$$\lim_{M \to \infty} M^{-k} \sum_m E[\epsilon_t \exp\{m' x_t(\delta) + \gamma' \Psi(x_t(\delta))\}]^2; \quad ((19'))$$

or, a la the sup-test of Andrews and Ploberger (1994), we may optimize over a countable vector subset $[0, M]^k \subset \mathbb{Z}^k$ of integers: $\lim_{M \to \infty} \sup_{m \in [0, M]} LM(m, \gamma, \delta)$.

Notice that $\exp\{m' x_t(\delta)\}$ may be identically represented as $\exp\{\delta' x_t(m)\}$ where $x_t(m) \equiv (m_1 x_{t,1}, ..., m_k x_{t,k})'$. Moreover, Theorem 4 holds for any $\gamma$, and therefore holds at the origin $\gamma = 0$. This implies there exist infinitely many super-consistent Bierens-type moment conditions of the form $E[\epsilon_t \exp\{\gamma' X_t(m)\}] \neq 0$ for every $\delta_i \neq 0$, $i = 1...k$. We summarize this result in the following corollary.

Corollary 5 Under the assumptions of Theorem 4, if $P[\epsilon_t | x_t] = 0 < 1$ then $E[\epsilon_t \exp\{\gamma' \Psi(x_t, m)\}] \neq 0$ for every $\gamma \in \Gamma$, $\gamma \neq 0$, for infinitely many bounded, one-to-one mappings $\Psi(x_t, m)$, in particular for $\Psi(x_t, m) = x_t(m)$, for infinitely many $m \in \mathbb{Z}_k^+$. 

Remark 1: As in Lemmas 1 and 3, we may replace $\epsilon_t$ with $\epsilon_t h(x_t, \delta)$, where $h(x_t, \delta)$ satisfies Assumption 1.

Remark 2: Following Remark 1 of Theorem 4, we may use $\Psi(x_t, m) = (m_1 \Psi_1(x_t), ..., m_k \Psi_k(x_t))'$ for any $\Psi(x_t, m)$ and $\Psi(x_t) : \mathbb{R}^k \to \mathbb{R}^k$ that satisfy Assumption A'.

Remark 3: Under $H_1$, Bierens (1990: Lemma 1) proves $E[\epsilon_t \exp\{\gamma' x_t\}] \neq 0$ except for a countable set $S$ of $\gamma$. We can now see this is a special case of Corollary 5 because the moment condition $E[\epsilon_t \exp\{\gamma' x_t\}]$ is simply a special case of $E[\epsilon_t \exp\{\gamma' x_t(m)\}]$ with fixed $m = (1, 1, ..., 1)'$. By allowing the integer vectors $m$ to take on any positive value, however, we can effectively fix $\gamma$ and search over $m$, a simpler task in practice. Of course, in practice it will be prudent to perform a two-dimensional search over various $\gamma$'s and $m$'s: see Section 6.

3.3 Conditioning Set $\Psi(x)$: Non-Degeneracy
Using Theorem 8 from the subsequent section and \( \Psi(x_t(\delta)) \), we can construct a moment condition that implies both LM statistic super-consistency and non-degeneracy: see Section 4.1 for general details. Consider the moment

\[
E \left[ \epsilon_t \partial / \partial \theta f(x_t, \theta_0) \exp \{ \delta' x_t(m) \} \prod_{i=1}^k F^{m_i-r_i} (\gamma' \Psi(x_t(\delta))) \right].
\]

The next result proves the asymptotic covariance matrix \( V(\gamma, \delta, m) \) associated with a sample version of the above moment is positive definite essentially for any \( \gamma, \delta, \) and \( m \). Define the set

\[
S^* = \{ \gamma \in \Gamma : q' V(\gamma, \delta, m) q = 0, \ q \in \mathbb{R}^k, \ q \neq 0 \}.
\]

**Theorem 6** Let the assumptions of Theorem 4 and Assumptions C and D in Section 4 hold. For any \( m \neq 0 \) the set \( S^* \) is empty; and for \( m = 0 \) the set \( S^* = \{0\} \).

**Example 3: ESTAR** Consider a test of linear autoregression in which \( \partial / \partial \theta f(x_t, \theta) = x_t \), consider an exponential weight \( F(u) = \exp(u) \), and define the argument \( \Psi(x_t, m, c) = (-m_1(x_{t,1} - c_1)^2, ..., -m_k(x_{t,k} - c_k)^2)' \), \( c_i \neq 0 \) for at least one \( i = 2...k \). Because Corollary 5 and Theorem 6 hold for \( \gamma = 0 \), if \( H_1 \) is true an optimal alternative model at \( \gamma = 0 \) and \( \delta_i > 0 \) can be derived as

\[
y_t = \theta' x_t + \beta' x_t \exp \{ \delta' \Psi(x_t, m, c) \} + \epsilon_t,
\]

an exponential smooth transition autoregression (ESTAR). A test with such an alternative in mind will be super-consistent and non-degenerate against any deviation from the null for any \( \delta_i > 0 \), \( i = 1...k \), and for infinitely many integer vectors \( m_i \in \mathbb{Z}^k \).

Bierens (1990: Corollary 1) shows the conditional expectation \( E[y_t|x_t] \) can be expressed as an infinite order series expansion in terms of exponential functionals. Using the above results, however, we can justify expressing \( E[y_t|x_t] \) as a linear combination of terms with any structure permitted by Lemma 3, \( h(x_t, \delta) F(\gamma' \Psi(x_t, m)) \).

**Corollary 7** Let \( y_t \) be a random variable satisfying \( E[y_t^2] < \infty \), and let \( x_t \) be an \( \mathfrak{S}_{k-1} \)-measurable vector in \( \mathbb{R}^k \). Define \( h(x_t, \delta) : \mathbb{R}^k \times \Delta \to \mathbb{R}^m \) as in Lemma 1. For any measurable mapping \( \Psi(x_t, m) = (m_1 \Psi_1(x_t), ..., m_k \Psi_k(x_t))' \) such that \( \Psi(x_t, m) \) and \( \Psi(x_t) \) satisfy Assumption \( A' \), any sequence of vectors \( \gamma_j \in \Gamma, \gamma_j \neq 0, \) and \( \delta_j \in \Delta, \) and infinitely many sequences of \( \mathbb{Z}^k \)-valued vectors \( m_i, i = 1, 2, ..., \) there exist \( k \)-vector coefficients \( \beta_{n,j}, j = 0, 1, 2, ..., \) with \( \beta_{n,0} \) a scalar for each \( n, \) such that

\[
E[y_t|x_t] = \beta_{0,0} + \sum_{n=1}^\infty \left[ \beta_{n,0} + \sum_{j=1}^n \beta_{n,j} h(x_t, \delta_j) F(\gamma_j' \Psi(x_t, m)) \right].
\]

**Remark 1:** Bierens (1990: Corollary 1) requires the set \( \gamma_1, \gamma_2, ... \), to be dense. The above result holds for any sequence of non-zero vectors \( \gamma_j \neq 0 \), as long as the integer vectors \( m_j \) are chosen (from an infinite integer set) appropriately.
Example 4 Let $F(u) = \exp(u)$, $h(x_i, \delta_j) = x_i$, and $\Psi(x_t, m, c) = (-m_1(x_{i,1} - c_1)^2, \ldots, -m_k(x_{i,k} - c_k)^2)'$, $c_i \neq 0$ for at least one $i$. Then the stipulations of Corollary 7 hold, and

\[ E[y_t|x_t] = \alpha_0 + \sum_{n=1}^{\infty} \sum_{j=1}^{n} \beta_{n,j} x_t \exp \left( -\sum_{i=1}^{k} \gamma_{j,i} m_{j,i} (x_{t,i} - c_i)^2 \right), \]

where $\alpha_0 = \beta_{0,0} + \sum_{n=1}^{\infty} \beta_{n,0}$. For any non-zero $\gamma_j$, and infinitely many $m_j$. Thus, $E[y_t|x_t]$ may be expressed as an infinite-regime ESTAR.

4. **Super-Consistency and Non-degeneracy for Arbitrary $\Psi(x_t)$** In Section 3.2 we considered the specific mapping $\Psi(x_t(\delta)) = (e^{x_{t,1}(\delta)}, \ldots, e^{x_{t,k}(\delta)})'$ and implications for super-consistency and non-degeneracy. In this section we generalize to any mapping $\Psi(x_t)$ that satisfies Assumption A', and develop moment condition vector weights $h(x_t, \delta)$ that variously imply non-degeneracy of an LM test, super-consistency, or both.

4.1 **Non-degeneracy of the Covariance Matrix** The LM test statistic based on (18) has the well-known form

\[ T_n(\gamma, \delta) = n \hat{s}(\gamma, \delta)' \hat{V}(\gamma, \delta)^{-1} \hat{s}(\gamma, \delta), \]

where $\hat{V}(\gamma, \delta)$ is any consistent estimator of $V(\gamma, \delta)$ in (11).

There exist cases in which $V(\gamma, \delta)$ is singular. For example, if the null specification is linear, $f(x_t, \theta) = \theta' x_t$, and if $h_t(\delta) \equiv x_t$ and $\gamma = 0$, then $T_t(\gamma)$ is constant-valued by Assumption 3 and $\hat{s}(\gamma, \delta) = 0$ by the least-squares first-order conditions, implying $V(\gamma, \delta) = 0$, a zero-matrix. For a consistent test statistic with non-degenerate limit distribution, we must therefore analyze the set of all $\gamma$ for which $V(\gamma, \delta)$ is non-positive definite.

Define the set

\[ S^* = \{ \gamma \in \Gamma : r' V(\gamma, \delta) r = 0, \ r \in \mathbb{R}^k, r \neq 0 \}. \]  

Denote by $\langle , \rangle$ the Euclidean inner-product: $\langle x, y \rangle = \sum_{i=1}^{k} x_i y_i$. Denote by $\mathcal{L}_{-1}^{k,\theta}(\theta_0)$ the space of all $k$-vector random variables $z_{t-1}$ that satisfy the following orthogonality condition: for every $z_t \in \mathcal{L}_{-1}^{k,\theta}(\theta_0)$, $\langle z_t, (\partial/\partial \theta) f(x_t, \theta_0) \rangle = 0$ with strictly positive probability.

**Assumption B** There exists a Borel-measurable real function $\mu$ on $\mathbb{R}^k$ such that the random vector $\kappa = (\mu(x_t), (\partial/\partial \theta) f(x_t, \theta_0))$ has nonsingular moment matrix $E[\kappa \kappa']$.

**Assumption C** For each $t$, $P(E[\kappa_t^2|x_t] > 0) = 1$ and $P(|x_t| > 0) = 1$.

**Assumption D** The matrix $b(\gamma, \delta)$ defined in (12) has full row and column rank uniformly in $\Gamma$ and $\Delta$. For every $z_t \in \mathcal{L}_{-1}^{k,\theta}(\theta_0)$, $z_t = 0$ with probability one.
Initially assume \( x_t \) does not contain a constant term, and let \( \gamma = (\gamma_0, ..., \gamma_{k-1})' \).

**Theorem 8** Assume there exists a matrix function \( g_t(\delta) \equiv g(x_t, \delta) : \mathbb{R}^k \times \Delta \rightarrow \mathbb{R}^{m_2 \times k} \) such that

\[
h(x_t, \delta) = g(x_t, \delta)\partial / \partial \theta f(x_t, \theta_0) .
\]

Let Assumptions 1-4 and Assumptions C and D hold. Then (i) if \( g(x_t, \delta) = g \) with probability one, where \( g \) is a constant, non-zero matrix, then \( S^* = \{0\} \); or (ii) if for every \( r \in \mathbb{R}^{m_2}, r \neq 0 \), \([r'g(x_t, \delta)]_l\) and \( F(\gamma'x_t) g_{i,j}(x_t, \delta) \) are \( \mathbb{R}_{i-1} \)-measurable with uniformly non-degenerate marginal distributions for every \( t \) and at least one \( l \) and one pair \( i, j \), then \( S^* \) is empty; or (iii) if \([r'g(x_t, \delta)]_l\) for some \( r \in \mathbb{R}^{m_2}, r \neq 0 \), has degenerate marginal distributions for every \( l = 1 \ldots k \) and \( F(\gamma'x_t) g_{i,j}(x_t, \delta) \) are \( \mathbb{R}_{i-1} \)-measurable with uniformly non-degenerate marginal distributions for every \( t \) and at least one pair \( i, j \), then \( S^* = \{0\} \).

**Remark 1:** Using only the scalar-weight \( F(\gamma'x_t) \), Bierens (1990) and de Jong (1996) require Assumptions B and C in order to prove \( S^* \) is countable, has Lebesgue measure zero and is nowhere dense: see Theorem 11, below, for a similar result. In the present context, however, under Assumptions C and D we obtain the result that the set \( S^* \) is either empty or contains only the origin. Assumption C is non-intrusive, and Assumption D restricts the gradient components \((\partial / \partial \theta) f(x_t, \theta_0)\) and columns of \( h(x_t, \delta)\partial / \partial \theta f(x_t, \theta_0)\) to be linearly independent.

**Remark 2:** Under cases (i) and (iii), if \( x_t \) contains a constant term then \( \gamma = 0 \) is only one element of \( S^* \): the "intercept" \( \gamma_0 \) need not be zero. For example, assume \( x_t = 1 \), denote by \( \gamma_{(i)} \) the \( k-1 \) sub-vector of \( \gamma \) with \( \gamma_i \) removed. Thus, \( \gamma_{(0)} = (\gamma_1, ..., \gamma_{k-1})' \), and denote by \( S^*_{(0)} \) and \( \Gamma_{(0)} \) the relevant sets associated with \( \gamma_{(0)} \):

\[
S^*_{(0)} = \{ \gamma_{(0)} \in \Gamma_{(0)} \subseteq \mathbb{R}^{k-1} : \sup_{\delta \in \Delta} |r'V(\gamma, \delta)r| = 0, \ r \in \mathbb{R}^k, \ r \neq 0 \}. \tag{28}
\]

Then \( S^*_{(0)} = \{0\} \) follows from the line of proof of Theorem 7, and we deduce \( S^* = \{ \gamma = (\gamma_0, \gamma_{(0)}) = (\omega, 0) : \omega \in \mathbb{R}, 0 \in \mathbb{R}^{k-1} \}. \)

**Remark 3:** Cases (ii) and (iii) place restrictions on \( F(\gamma'x_t) g_{i,j}(x_t, \delta) \) similar in spirit to the restrictions imposed on \( h(x_t, \delta)F(\gamma'x_t) \) in Lemmas 1 and 3. The present restrictions merely imply we cannot define, for example, \( g(x_t, \delta) = I_k \times F_t(\gamma)^{-1} \), for \( F_t(\gamma) \neq 0 \) with probability one: in such a case \( h(x_t, \delta)F_t(\gamma) = g(x_t, \delta)\partial / \partial \theta f(x_t, \theta_0)F_t(\gamma) = I_k \times F_t(\gamma)^{-1} \partial / \partial \theta f(x_t, \theta_0)F_t(\gamma) = \partial / \partial \theta f(x_t, \theta_0)\), and the sample score would satisfy \( \hat{s}(\gamma, \delta) = 0 \) by the nonlinear least-squares first order conditions, a conditional moment failure under \( H_1 \), and \( V(\gamma, \delta) \) would be degenerate.

Even Assumption D, however, can be simplified in certain contexts. In particular, if \( \partial / \partial \theta f(x_t, \theta_0) = x_t \) and \( g(x_t, \delta) = g \), a constant non-zero matrix, then we may omit the additional orthogonality assumption: \( \langle z_t, x_t \rangle = 0 \) with strictly positive probability if and only if \( z_t = 0 \) with probability one now holds by Assumption 1. We summarize this in the following corollary.
Corollary 9 Under the assumptions of Theorem 7, if $\partial \partial_f (x_t, \theta) = x_t$, and $g(x_t, \delta) = g$ a constant non-zero matrix, then $S^* = \{(\omega, 0) : \omega \in \mathbb{R}, 0 \in \mathbb{R}^{k-1}\}$, assuming a constant term is included, and $S^* = \{0\}$ otherwise.

Example 5 The moment condition for a test of linear autoregression against a smooth transition autoregression (STAR) exactly satisfies the functional requirements of Corollary 9. The linear null models has $f(x_t, \theta) = \theta'x_t$, and a smooth transition-type alternative is represented as

$$y_t = \theta'x_t + \beta x_tF(\gamma'\Psi(x_t)) + \epsilon_t \quad (29)$$

for either an exponential (ESTAR) or logistic (LSTAR) functional $F(\gamma'\Psi(x_t))$. See, e.g., Teräsvirta (1994). Thus, $h(x_t, \delta) = g(x_t, \delta)\partial \partial_f (x_t, \theta_0) = 1 \times x_t$, hence an LM test of linearity against smooth transition nonlinearity is asymptotically non-degenerate. Moreover, if we use $F(\gamma'x_t(m))$ the test will be super-consistent for infinitely many $m \in \mathbb{Z}_+$. See Example 1 for issues concerning the weight argument $\Psi(x_t)$ for ESTAR models.

4.2 Super-Consistency Denote by $\Gamma$ a bounded compact subset of $\mathbb{R}^k$. Let $\Gamma$ be partitioned into dimension-specific scalar-spaces: $\Gamma = \prod_{i=1}^k \Gamma_i$, where $\Gamma_i \subset \mathbb{R}_+$ (the weakly positive hyper-sphere). Consider an arbitrary function $F_\gamma(\gamma) = F(\gamma'\Psi(x_t))$ that satisfies Assumption $A'$. Define the moment

$$\eta^{(1)}(\gamma) \equiv E[\epsilon_tF(\gamma'\Psi(x_t))], \quad (30)$$

and define the parameter matrix $\gamma^{(*)} = (\gamma^{(1)}, \ldots, \gamma^{(k)})$, where $\gamma^{(i)} \in \Gamma_i$ is a $k$-vector for each $i$, as follows:

$$\gamma^{(i)} = \arg \sup_{\gamma \in \Gamma_i} \left\{ (\partial / \partial \gamma_i)\eta^{(1)}(\gamma) \right\}. \quad (31)$$

Thus, for each $i$, $\gamma^{(i)}$ generates the largest slope:

$$(\partial / \partial \gamma_i)\eta^{(1)}(\gamma) - (\partial / \partial \gamma_i)\eta^{(1)}(\gamma)|_{\gamma = \gamma^{(i)}} \leq 0. \quad (32)$$

For the sake of convention, assume $(\partial / \partial \gamma_i)\eta^{(1)}(\gamma)|_{\gamma = \gamma^{(i)}} \geq 0$.

In general, the set $\gamma^{(*)}$ need not be unique under either hypothesis. For example, under the null hypothesis the moment condition $(\partial / \partial \gamma)\eta^{(1)}(\gamma) = E[\epsilon_t\Psi(x_t)F_\gamma(\gamma)] = 0$ holds with probability one for any $\gamma$ by the assumed martingale difference property of the innovations series $\epsilon_t$. Thus, strict equality in (32) holds for any $\gamma$ and $\gamma^{(i)}$ under the $H_0$. Similarly, under $H_1$, provided $\Psi(x_t)$ is bounded and non-zero with probability one and $\Psi(x_t)F_\gamma(\gamma)$ satisfies Assumption A, we deduce the moment condition $(\partial / \partial \gamma)\eta^{(1)}(\gamma) = E[\epsilon_t\Psi(x_t)F_\gamma(\gamma)] \neq 0$ holds except for a countable set of $\gamma$, hence $(\partial / \partial \gamma)\eta^{(1)}(\gamma) = E[\epsilon_t\Psi(x_t)F_\gamma(\gamma)] \neq 0$ for uncountably many $\gamma \in \Gamma$, and therefore $(\partial / \partial \gamma_i)\eta^{(1)}(\gamma)|_{\gamma = \gamma^{(i)}} \neq 0$. However, the argument $\gamma^{(i)} = \arg \sup(\partial / \partial \gamma_i)\eta^{(1)}(\gamma)|_{\gamma \in \Gamma}$ need not be unique, and may be zero. If $\gamma^{(*)}$ is not unique, we simply write $\gamma^{(*)}$ to denote any matrix that satisfies (32).
For arbitrary $\gamma \in \Gamma$ define the $k + 1$-vector weight $h_i^{(1)}(\gamma, \gamma^*)$ as

$$h_i^{(1)}(\gamma, \gamma^*) = [1, \Psi_1(x_i)F_i'((\gamma^{(i)}))/F_i(\gamma), ..., \Psi_k(x_i)F_i'((\gamma^{(k)}))/F_i(\gamma)]^f$$

(33)

where $F'(u) = (\partial/\partial u)F(u)$, and $F_i(\gamma) \neq 0$ with probability one by Assumption 3. For simplicity, assume $x_t$ does not contain a constant term.

**Theorem 10** Let $\epsilon$ be a random variable satisfying $E|\epsilon_i| < \infty$, and let $x_t$ be an $\mathcal{F}$-measurable bounded vector in $\mathbb{R}^k$, $0 < k < \infty$, such that $P[E(\epsilon_i|x_t) = 0] < 1$. Assume $\Gamma$ is a bounded subset of $\mathbb{R}^k$, and let Assumption $A'$ hold for $F(\gamma^*\Psi(x_t))$ and $h_i^{(1)}(\gamma, \gamma^*)F(\gamma^*\Psi(x_t))$. Then each set

$$S_i^{(1)} = \{ \gamma \in \Gamma : E[\epsilon_i h_i^{(1)}(\gamma, \gamma^*)F(\gamma^*\Psi(x_t))] = 0, \text{ and } P(\gamma^*\Psi(x_t) \in R_0) = 1 \}$$

contains only the singleton $\gamma = 0$ if and only if $\epsilon_t \perp \mathcal{F}|[\Psi_1(x_t)F_i'(\gamma^{(i)})]$, and $S_i^{(1)} = \emptyset$ otherwise.

**Remark 1:** The result remains valid if we restrict $\Gamma_i$ to be any bounded sub-set of the negative hyper-sphere $\mathbb{R}^k$.

**Remark 2:** If $\gamma^{(i)} = 0$ for each $i = 1...k$, then $F_i'(\gamma^{(i)}) = F_i'(0) = \hat{c}$ for some scalar constant $\hat{c}$, cf. Assumption 3. In this case $\gamma = 0$ is an element of $S_i^{(1)}$ if and only if $E[\epsilon_i\Psi(x_t)] = 0$. For example, if $\Psi(x_t) = x_t$ then $E[\epsilon_i\Psi(x_t)] = 0$ is automatically satisfied by Assumption 1, hence $S_i^{(1)} = \{0\}$.

**Remark 3:** If $x_t$ contains a constant term, then, using logic identical to arguments following Theorem 8, we conclude each $S_i^{(1)} = \{ \gamma = (\omega, 0): \omega \in \mathbb{R}, 0 \in \mathbb{R}^{k-1} \}$.

**Remark 4:** Rather than assume $h_i^{(1)}(\gamma, \gamma^*)$ and $F_i(\gamma)$ are functions of the same vector $\gamma$, write $h_i^{(1)}(\hat{\gamma}, \gamma^*)$ and $F_i(\hat{\gamma})$ for some point $\hat{\gamma} \in \Gamma$. Note that $h_i^{(1)}(\hat{\gamma}, \gamma^*)$ is uniformly bounded and non-zero with probability one: see the line of proof. Then the sets

$$S_{\hat{\gamma}, i}^{(1)} = \{ \gamma \in \Gamma : E[\epsilon_i h_i^{(1)}(\hat{\gamma}, \gamma^*)F_i(\hat{\gamma})] = 0, \text{ and } P(\gamma^*x \in R_0) = 1 \},$$

(35)

have Lebesgue measure zero by Lemma 3, however the contents of $S_{\hat{\gamma}, i}^{(1)}$ may involve more than the singleton $\{0\}$. Theorem 10, however, implies the element $\hat{\gamma}$ itself is not contained in $S_{\hat{\gamma}, i}^{(1)}$. Indeed, the set $S_i^{(1)}$ is identically the union of all planes of $S_{\hat{\gamma}, i}^{(1)}$ where $\hat{\gamma} = \gamma$, which we denote as $S_i^{(1)}(\hat{\gamma})$: hence $S_i^{(1)}(\hat{\gamma}) = \bigcup_{\hat{\gamma} \in \Gamma} S_{\hat{\gamma}, i}^{(1)}(\hat{\gamma})$. Thus, projecting the vector moment condition $E[\epsilon_i h_i^{(1)}(\hat{\gamma}, \gamma^*)F_i(\hat{\gamma})]$, with support $\Gamma \times \Gamma$, onto the plane where $\hat{\gamma} = \gamma$ promotes super-consistency.

**Remark 5:** The above moment condition $E[\epsilon_i h_i^{(1)}(\gamma, \gamma^*)F_i(\gamma)]$ provides a two-way safety net against failure of Bierens test consistency under the alternative. If under $H_1$ a chosen vector $\gamma$ implies a failure of the Bierens-type moment condition such that

$$E[\epsilon_i F_i(\gamma)] = 0,$$

(36)
then, for an appropriately defined mapping $\Psi(x_t)$, the condition
\[ E \left[ \epsilon_t \Psi_t(x_t) F'_t(\gamma^{(i)}) \right] \neq 0 \]  \hspace{1cm} (37)

is guaranteed to hold for at least one $i \in \{1, ..., k\}$. Conversely, because the vectors $\gamma^{(i)}$ only maximize each gradient element $E[\epsilon_t \Psi_t(x_t) F'_t(\gamma^{(i)})]$ (and not the absolute magnitude), it is still possible that $E[\epsilon_t \Psi_t(x_t) F'_t(\gamma^{(i)})] = 0$ for each $i = 1..k$. In such a case, however, the moment condition $E[\epsilon_t \exp(\gamma' x_t)] \neq 0$ is guaranteed to hold. Each moment condition fills in the break-down points of the other, and together provide a super-consistent parametric test framework.

**Remark 6:** The theory detailed in Section 3 can be directly applied to the present setting. For example, using Corollary 5 we can consider the moment condition $\eta^{(i)}(\gamma, m) \equiv E[\epsilon_t \exp(\gamma' x_t(m))]$ for $m \in \mathbb{Z}^k$. In this case, non-zero $\gamma$ is arbitrary, we maximize $\eta^{(i)}(\gamma, m)$ with respect to $m$, and therefore $E[\epsilon_t h^{(1)}_t(\gamma, m^{(\gamma)}) \exp(\gamma' x_t(m))] \neq 0$ under $H_1$ for infinitely many $m$ and any $\gamma \neq 0$, where
\[ h^{(1)}_t(\gamma, \gamma^{(i)}) = (1, m_1 x_{t,1,1} \exp(\gamma' \{ x_t(m^{(1)}) - x_t(m) \}), ..., m_k x_{t,k} \exp(\gamma' \{ x_t(m^{(k)}) - x_t(m) \}))' \] \hspace{1cm} (38')

Although the asymptotic covariance matrix cannot be proven to be positive definite everywhere in $\Gamma$, it can be shown that the set $S^{(1)}$ of nuisance vectors for which the covariance matrix is non-positive definite is countable.

**Theorem 11** Under Assumptions B and C the set $S^{(1)}$ has Lebesgue measure zero and is nowhere dense.

### 4.3 Super-Consistency and Non-degeneracy

Again, denote by $\Gamma$ a bounded compact subset of $\mathbb{R}^k_+$ and partition as before $\Gamma = \prod_{i = 1}^k \Gamma_i$, where $\Gamma_i \subset \mathbb{R}_+$. Define the moment
\[ \eta^{(2)}(\gamma) = E[\epsilon_t \Psi_t(x_t) F_1(\gamma)]. \]  \hspace{1cm} (38)

Define the nuisance parameter $k$-vectors $\gamma^{(i,j)} \in \Gamma \subseteq \mathbb{R}^k$, $i, j = 1..k$, as follows
\[ \gamma^{(i,j)} = \arg \sup_{\gamma \in \Gamma} \{ (\partial/\partial \gamma_j) \eta^{(2)}_t(\gamma) \}. \hspace{1cm} (39) \]

Because $\partial/\partial \gamma_j \eta^{(2)}(\gamma) = E[\epsilon_t \Psi_t(x_t) \Psi_j(x_t) F'_1(\gamma)] = \partial/\partial \gamma_j \eta^{(2)}_t(\gamma)$, symmetry follows $\gamma^{(i,j)} = \gamma^{(j,i)}$. Using the same arguments put forth in Section 4.2, the nuisance parameter set $\gamma^{(i,j)}$ need not be unique under either null or alternative hypotheses.

Denote by $\gamma^{(s)}$ the set of $k$-vectors $\gamma^{(i,j)}$, $i, j = 1..k$, $\gamma^{(s)} = (\gamma^{(i,j)})_{i,j=1}^k$. For arbitrary $\gamma \in \Gamma$ and for each $i, j = 1..k$ define the weights
\[ h^{(2)}_{i,i,j}(\gamma, \gamma^{(s)}) = \Psi_t(x_t) \Psi_j(x_t) F'_1(\gamma^{(i,j)} x_t)/F(\gamma' x_t), \hspace{1cm} (40) \]

and denote by $h^{(2)}_t(\gamma, \gamma^{(s)})$ the $m_2 = k(5 + k)/2$-vector that vertically stacks all the vector-weights $\Psi(x_t) F'_1(0)/F(\gamma)$, $\Psi(x_t)$ and $h^{(2)}_{i,i,j}(\gamma)$, $1 \leq i \leq j \leq k$. Again, for simplicity assume $x_t$ does not contain a constant term.
Theorem 12 Let $c$ be a random variable satisfying $E|c| < \infty$, and let $x$ be an $\mathbb{F}$-measurable bounded vector in $\mathbb{R}^k$, $0 < k < \infty$, such that $P[E(c|x) = 0] < 1$. Assume $\Gamma$ is a bounded subset of $\mathbb{R}^k$, and let Assumption $A'$ hold for $F(\gamma'\Psi(x_t))$ and $h^{(2)}_t(\gamma,\gamma^{(s)})F(\gamma'\Psi(x_t))$. Then the set

$$S^{(2)}_i = \{ \gamma \in \Gamma : E[h^{(2)}_i(\gamma,\gamma^{(s)})F(\gamma'x_t)] = 0, \text{ and } P(\gamma'\Psi(x_t) \in R_0) = 1 \},$$

contains only the singleton $\gamma = 0$ if and only if $c_t \perp \psi_t \left[ \psi_i(x_t)\psi_j(x_t)F(\gamma^{(i,j)}) \right]$, and $S^{(2)}_i = \emptyset$ otherwise.

Remark 1: If $x_t$ contains a constant term, then, using logic identical to arguments following Theorem 8, we conclude $S^{(2)} = \{ \gamma = (\omega,0) : \omega \in \mathbb{R}, 0 \in \mathbb{R}^{k-1} \}$.

The above result proves the weight $h^{(2)}_t(\gamma,\gamma^{(s)})$ promotes super-consistency. For non-degeneracy define the set

$$S^{(2)*} = \{ \gamma \in \Gamma : r'V(\gamma,\delta)r = 0, \ r \in \mathbb{R}^{m_2}, r \neq 0 \}$$

where the covariance matrix $V(\gamma,\delta)$ is defined by (11) and (12) with $h(x_t,\delta)$ replaced by $h_t^{(2)}(\gamma,\gamma^{(s)})$, where $\delta = \gamma^{(s)}$.

Corollary 13 Assume $f(x_t,\theta) = \theta'x_t$ and $\Psi(x_t) = x_t$. Under Assumptions C and D the set $S^{(2)*} = \{ 0 \}$.

Remark 1: There is much flexibility regarding the linear restriction on the null specification $f(x_t,\theta)$, depending on the construction of the weight $h^{(2)}_t(\gamma,\gamma^{(s)})$ and conditioning vector $\Psi(x_t)$. Indeed, for chosen null specification $f(x_t,\theta)$, we can simply set $\Psi_i(x_t) = (\partial\partial\theta_i)f(x_t,\theta)$ as long as Assumption $A'$ holds. For example, in the exponential regression model set $f(x_t,\theta) = \exp(\theta'x_t)$, put $\Psi_i(x_t) = \exp(\theta'x_t)$, and consider the exponential weight $F(u) = \exp(u)$. In this case, $h^{(2)}_t(\gamma,\gamma^{(s)})$ stacks $x_t \exp((\theta - \gamma')x_t, x_t \exp(\theta'x_t)$, and $x_{t,i}\exp((2\theta + \gamma^{(i,j)} - \gamma')x_t)$.

Remark 2: If $x_t$ contains a constant term, then, using logic identical to arguments following Theorem 8, we conclude $S^{(2)*} = \{ \gamma = (\omega,0) : \omega \in \mathbb{R}, 0 \in \mathbb{R}^{k-1} \}$.

Together, Theorem 12 and Corollary 13 imply an LM test based on the sample moment condition

$$\hat{s}(\gamma,\delta) = \frac{1}{n} \sum_{t=1}^{n} (y_t - f(x_t,\hat{\theta}))h^{(2)}_t(\gamma,\gamma^{(s)})F(\gamma'x_t)$$

will be consistent and non-degenerate for any nuisance vector $\gamma$ in any compact subset of $\mathbb{R}^k$, except $\gamma = (\omega,0) : 0 \in \mathbb{R}^{k-1}$, assuming a constant term is used.

5. Tests of Linearity against Logistic Alternatives Use of the exponential and logistic as "totally revealing" functional forms in the neural networks and smooth transition autoregression literatures has expanded dramatically over
the last two decades: see, e.g., Kuan and White (1994). Although such weights can be used to construct super-consistent and/or non-degenerate tests of functional mis-specification, they also provide popularly recognized empirical interpretations that are interesting in their own right. In this section, we exemplify the various alternative models implied by the moment conditions discussed in Section 4 for these popular functional weights. By far the clearest examples exist for the case of tests of linearity against general alternatives, and in the case where \( \Psi(x_t) = x_t \). In the following, consider \( f(x_t, \theta) = \theta' x_t, \Psi(x_t) = x_t \), and logistic \( F(u) \), and consider the weights of Section 4.

Let \( F_t(\gamma) = (1 + \exp(\gamma_0 x_t))^{-1} \). The following models are implied under the alternative of functional mis-specification, for (i) non-degenerate, (ii) super-consistent, or (iii) both, respectively:

\[
(i) \quad y_t = \theta' x_t + \beta' x_t (1 + e^{\gamma' x_t})^{-1} + \epsilon_t; \\
(ii) \quad y_t = \theta' x_t + \beta (1 + e^{\gamma' x_t})^{-1} + \sum_{i=1}^{k} \psi_{t,i} x_t (1 + e^{\gamma^{(i)} x_t})^{-1} + \epsilon_t \\
(iii) \quad y_t = \theta' x_t + \beta' x_t (1 + e^{\gamma' x_t})^{-1} \\
\quad \quad + \sum_{1 \leq i \leq j \leq k} \psi_{t,i,j} x_t, x_t, x_t (1 + e^{\gamma^{(i,j)} x_t})^{-1} + \epsilon_t.
\]

The alternative functional form for (i) is simply an LSTAR structure. Under (ii) the alternative model has stochastic slopes defined by the smooth transition type process \( \psi_{t,i} = \psi_i (1 + \exp(-\gamma^{(i)} x_t))^{-1} \). Finally, under (iii) we obtain a model with random coefficients and smooth transition structures, where the stochastic coefficients are identically \( \psi_{t,i,j} = \psi_{t,i,j} (1 + \exp(-\gamma^{(i,j)} x_t))^{-1} \). Thus, in a logistic framework the class of alternative models that induces a functional form mis-specification test that is both super-consistent and non-degenerate implies an extremely versatile functional form with functional regimes that incrementally augments the depth of tested nonlinearity: the first regime is linear \( \theta' x_t \); the second a standard LSTAR form \( \beta' x_t (1 + \exp(\gamma' x_t))^{-1} \); the third a cross-product amalgamation with smooth transition dynamics unique to each product \( x_t x_t \) and coefficients \( \psi_{t,i,j} \) sensitive to continuous threshold dynamics.

6. Simulation Study We now investigate the empirical size and power properties of sup-LM tests of linear autoregression against LSTAR, ESTAR and bilinear alternatives.
6.1 Set-up  Our simulations are based on the following models:

\[
H_0 : y_t = \theta' x_t + \epsilon_t \\
H_1^R : y_t = \theta' x_t + \theta_2(t) x_t + \epsilon_t \\
H_1^S : y_t = \theta_1 x_t + \theta_2^t x_t I(y_{t-1} > c) + \epsilon_t \\
H_1^L : y_t = \theta_1 x_t + \theta_2 x_t (1 + \exp(-\gamma y_{t-1}))^{-1} + \epsilon_t \\
H_1^{BL1} : y_t = \theta_1 x_t + \theta_2 y_{t-1} \epsilon_{t-1} + \epsilon_t, \ |\theta_2| < 1 \\
H_1^{BL2} : y_t = \theta_1 x_t + y_{t-1} \epsilon_{t-1} + \epsilon_t \\
H_1^{BL3} : y_t = \theta_1 y_{t-1} \epsilon_{t-1} + \epsilon_t, \ |\theta_1| < 1,
\]

where \(\epsilon_t\) are iid standard normal, and \(x_t = (1, y_{t-1}, ..., y_{t-p})'\) for some \(p \geq 0\).

Under \(H_0\) the true data generating process is linear; under \(H_1^R\) the linear process is driven by a white-noise parameter set \(\theta_2(t)\) uniformly distributed on the hypercube \([-0.9, 0.9]^k\); under \(H_1^S\) the process is a standard 2-regime self-exciting autoregression (SETAR); under \(H_1^L\) the true process is a 2-regime logistic STAR; and under each \(H_1^{BL}\), the process is bilinear. We consider sample sizes \(n = 100, 500,\) and \(1000\): in each case, we generate \(3n\) observations, and retain the last \(n\) in order to reduce dependence on starting values.

For each simulated series, the order \(p\) is randomly chosen from the set \(\{1, ..., 10\}\), and the vectors \(\theta_t\) are randomly chosen from the hypercube \([-1.5, 1.5]^k\). For SETAR processes, the threshold \(c\) is randomly selected from the uniform interval \([-2, 2]\). Because we assume the process \(y_t\) is covariance stationary under the null, only vectors \(\theta_1\) with characteristic polynomial roots outside the unit circle are considered. For STAR processes, the scale \(\gamma\) is randomly drawn from the uniform interval \([0.1, 10]\).

We generate 1000 replications of each series above, a linear model is estimated and the resulting residuals are tested. In order to specify the linear autoregressive null model, we employ a minimum AIC selection criterion for the order \(p\). For model selection, we always include an intercept. A residual series \(\{\hat{\epsilon}_t\}_{t=1}^n\) is generated and the following tests are performed at the 5%-level.

6.2 Tests of Functional Form  We perform non-degenerate, super-consistent and super-consistent/non-degenerate versions of the functional specification tests implied by the moment conditions of Sections 3 and 4. We also perform the standard Bierens (1990) sup-test and the neural test of neglected nonlinearity, cf. Lee, White and Granger (1996). All tests are performed as score tests, thus denote the null score as \(\hat{s}(\gamma) = n^{-1} \sum_{t=1}^n \hat{\epsilon}_t z_t(\gamma, m)\) where \(z_t(\gamma, m)\) denotes a set of weights implied by each test. For simplicity, because real \(\gamma\) and integer-valued \(m\) are undefined under the null hypothesis, in order to compute the test statistics we consider sup-LM tests, cf. Davies (1977) and Andrews and Ploberger (1994).

Super-Consistent, Non-degenerate Tests  
From Section 3, a super-consistent and non-degenerate test can be constructed for any \(\gamma \neq 0\). For a score test, the implied weight is \(z_t(\gamma, m) = x_t \exp\{\gamma' x_t(m)\}\). Denote by \(T_n^{(c)}(\gamma, m)\) the associated LM statistic.
From Section 4, super-consistent and/or non-degenerate tests are available. Fixing $m$ to be a column-vector of ones, for the non-degenerate test, the score weight is $z_t(\gamma,1) = x_t F(\gamma'x_t)$: in all cases, we use $\Psi(x_t) = x_t$. For the present and all subsequent tests, we consider both exponential and logistic weights $F(u)$.

Denote by $T_n^{(nd)}(\gamma)$ the associated test statistic. For the super-consistent test, denoted $T_n^{(sc)}(\gamma)$, the implied weight is $z_t(\gamma,1) = h_t^{(1)}(\gamma,\gamma^*) F(\gamma'x_t)$. Finally, for the non-degenerate, super-consistent test, denoted $T_n^{(b)}(\gamma)$, the implied weight is $z_t(\gamma,1) = h_t^{(2)}(\gamma,\gamma^*) F(\gamma'x_t)$.

Extensive simulation experimentation not presented here clearly demonstrates that for the moment conditions of Section 4 the columns of $z_t$ suffer from near perfect multicollinearity in some cases depending on $x_t$, the optimally selected $\gamma$, $\gamma(i)$, and $\gamma(i,j)$, leading to empirical sizes 2-5 times as large as nominal values.

We employ a simple, although inefficient, remedy which breaks the vector moment conditions apart into the separate scalar moments and optimizes over the associated individual test statistics. By Theorem 10, the set of $\gamma$ such that $E[e_t h_t^{(1)}(\gamma,\gamma^*) F(\gamma'x_t)] = 0$ contains only the trivial vector $\gamma = 0$, hence at least one moment condition $E[e_t h_t^{(1)}(\gamma,\gamma^*) F(\gamma'x_t)]$ is guaranteed to be non-zero, and therefore at least one LM test statistic based on the separate $k + 1$ sample scores $n^{-1} \sum_{t=1}^n e_t h_t^{(1)}(\gamma,\gamma^*) F(\gamma'x_t)$ will be consistent for any $\gamma \neq 0$. Hence, we need only derive each of the implied $k + 1$-test statistics separately, denoted $T_n^{(sp)}(\gamma)$, $i = 1...k$, and for each $\gamma$ choose the largest of the $k + 1$-statistics, denoted $T_n^{(sp)}(\gamma) = \max_{1 \leq i \leq k+1} \{T_n^{(sp)}(\gamma)\}$. By identical reasoning, we may employ $T_n^{(b)}(\gamma) = \max_{1 \leq i \leq k+1} \{T_n^{(b)}(\gamma)\}$ for the super-consistent, non-degenerate test. Asymptotically, for any $\gamma \neq 0$ the statistics $T_n^{(1)}(\gamma)$ and $T_n^{(2)}(\gamma)$ are super consistent.

For each test $T_n^{(sp)}(\gamma)$ and $T_n^{(b)}(\gamma)$ the vectors $\gamma(i)$ and $\gamma(i,j)$ are selected by respectively maximizing the sample moments $n^{-1} \sum_{t=1}^n e_t x_{ti} F_t'(\gamma)$ and $n^{-1} \sum_{t=1}^n e_t x_{ti} x_{tj} F_t'(\gamma)$ over the subset $\Gamma = [1,10]^k$, with increments set to .01.

Once the vectors $\gamma(i)$ and $\gamma(i,j)$ are selected, we maximize each $T_n^{(i)}(\gamma)$, $i \in \{ng, sc, b\}$ over the support $\Gamma$, and denote the final sup-LM statistics as $g_{n(i)} = \sup_{\gamma \in \Gamma} T_n^{(i)}(\gamma)$. Similarly, we maximize $T_n^{(m)}(\gamma, m)$ over $\gamma$ in $\Gamma$ and $m$ in the the set of integers $\{0,1,...,\sqrt{m}\}$, with $[w]$ the integer part of $w$.

With the statistics $g_{n(i)}$ in hand, $i \in \{e, nd, sc, b\}$, we employ Hansen’s (1996) parametric bootstrap method for approximating the asymptotic $p$-values. For Hansen’s method we simulate $J$ iid standard normal random $n$-vectors $(u_{t,j})_{t=1}^n$, $j = 1...J$, generate $J$ scores $s_{n,j}(\gamma) = n^{-1} \sum_{t=1}^n e_t u_{t,j} z_t(\gamma)$, $J$ test statistics, $T_{n,j}^{(i)}(\gamma)$, and $J$ statistic functionals $g_{n,j}^{(i)}$. The $p$-value is the percent frequency of the event $g_{n,j}^{(i)} > g_{n(i)}$. Under our Assumptions 1-4, Hansen’s (1996) Assumptions 1-3 and Theorems 1-2 hold\(^4\), and therefore the approximate $p$-value converges in

\(^4\)Specifically, Hansen’s (1996) Assumption 1 defines the process $\{y_t, x_t\}$ as strictly sta-
probability to the true $p$-value. For all simulations, we set $J = 500$. Together, these are the CoSTAR$^{(i)}$ tests.

**Bierens Test, Neural Test, etc.**

For the Bierens test, denoted BIER, $z_t(\gamma, 1) = F(\gamma'x_t)$. We maximize the LM statistic over the set $\Gamma$, and compute the $p$-value by Hansen’s (1996) bootstrap technique. In order to generate a test statistic directly comparable to every other test performed here, we by-pass the criterion technique detailed in Bierens (1990) which generates an asymptotically chi-squared statistic. Similarly, we perform the neural test of neglected nonlinearity, cf. Lee *et al* (1996), which is equivalent to a randomized Bierens test over the nuisance parameter space: $\gamma$ is selected randomly from the set $\Gamma$, and $p$-values are computed from the asymptotic $\chi^2(1)$-distribution. This is the NEURAL test.

For all LM tests employed in this study, covariance matrix estimators robust to unknown forms of conditional heteroscedasticity are used.

6.3 Results Results for $H_0$ are contained in Table 1, and Tables 2-4 contain empirical powers for the various alternatives. See Appendix 2.

**[UNDER CONSTRUCTION]**

Linear Autoregression
STAR
Bilinear and Random Coefficient

7. Conclusion
Appendix 1: Assumptions

Assumptions 1, 2 and 4 are standard for consistency and asymptotic normality of the nonlinear least square estimator: see Jenrich (1969) and White (1982).

**Assumption 1** The data-generating process \{y_t, x_t\} is defined in \(L_2(\Omega, P, \mathcal{F}_t)\) where \(\mathcal{F}_t\) denotes a strictly increasing \(\sigma\)-algebra induced by \((y_{t-1}, x_{t-1}), i = 0, 1, \ldots,\) such that \(\mathcal{F}_{t-1} \subset \mathcal{F}_t\). The regressors \(x_t\) are bounded \(k\)-vectors, measurable with respect to \(\mathcal{F}_{t-1}\). The process \{\(y_t, x_t\)\} is strictly stationary, ergodic, governed by non-degenerate joint distribution function with non-degenerate marginal distributions, and for some \(r > 1\), \(E|y_t|^r < \infty\). The innovations \(e_t\) form a \((0, \sigma^2)\)-martingale difference sequence with respect to \(\mathcal{F}_{t-1}\) under \(H_0\). Under \(H_1\), \(e_t\) is a \((0, \sigma^2)\)-white noise process for each \(t\) such that \(E[e_t x_t] = 0\).

**Assumption 2** The parameter space \(\Theta\) is a compact, convex subset of \(\mathbb{R}^k\). The real function \(f(x_t, \theta)\) maps from \(\mathbb{R}^k \times \Theta\) to \(\mathbb{R}\), and for each \(k\)-vector, is a compact, convex subset of \(\mathbb{R}^m\) for some \((m_1, m_2)\in \mathbb{Z}_+\), \(0 < \inf_{\delta \in \Delta} \|h(x_t, \delta)\| \leq \sup_{\delta \in \Delta} \|h(x_t, \delta)\| < \infty\) with probability one. \(h(x_t, \delta)\) is uniformly twice continuously differentiable in \(\delta\). The function \(F(\gamma \Psi(x_t)) : \mathbb{R}^k \times \Gamma \to \mathbb{R}\) is \(\mathcal{F}_{t-1}\)-measurable, bounded, where \(\Gamma\) is a bounded subset of \(\mathbb{R}^k\), and \(\Psi(x_t)\) maps from \(\mathbb{R}^k\) to \(\mathbb{R}^k\). Moreover, \(F(\gamma \Psi(x_t))\) is non-zero with probability one for any \(x_t\), and is constant-valued with probability one only at \(\gamma = 0\) if \(x_t\) does not contain a constant, and at \(\gamma = (w, 0), w \in \mathbb{R}, 0 \in \mathbb{R}^{k-1}\) if \(x_t\) contains a constant. Furthermore, each \((\partial/\partial u)^*F(u)\) is bounded in probability: there exists some \(N < \infty\) such that \(\forall s \in \mathbb{N}, P(\|(d/du)^*F(u)\| < N) = 1\).

**Assumption 4** The following uniform moment bounds hold for each \(t\), each \(i, j = 1..., k\), and for any \(h(x_t, \delta)\) and \(F_t(\gamma)\) defined by Assumption 3:

\[
E[\sup_{\theta \in \Theta} |(y_t - f(x_t, \theta))^2 \partial/\partial \theta_i f(x_t, \theta) \partial/\partial \theta_j f(x_t, \theta)|] < \infty;
\]

\[
E[\sup_{\theta \in \Theta} |(y_t - f(x_t, \theta)) \partial^2/\partial \theta_i \partial \theta_j f(x_t, \theta)|] < \infty;
\]

\[
E[\sup_{\theta \in \Theta} |\partial/\partial \theta_i f(x_t, \theta) \partial/\partial \theta_j f(x_t, \theta)|] < \infty.
\]

Moreover, \(A = E [\partial/\partial \theta^T f(x_t, \theta) \times \partial/\partial \theta^T f(x_t, \theta)]\) is positive definite uniformly in \(\Theta\).
## Appendix 2: Tables

### Table 1

<table>
<thead>
<tr>
<th>H₀ : linear AR</th>
<th>100</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>CoSTAR⁵ L</td>
<td>.127</td>
<td>.041</td>
<td>.035</td>
</tr>
<tr>
<td>CoSTAR⁵ E</td>
<td>.134</td>
<td>.012</td>
<td>.002</td>
</tr>
<tr>
<td>CoSTAR⁷ L</td>
<td>.087</td>
<td>.051</td>
<td>.032</td>
</tr>
<tr>
<td>CoSTAR⁷ E</td>
<td>.154</td>
<td>.042</td>
<td>.029</td>
</tr>
<tr>
<td>CoSTAR⁸ L</td>
<td>.042</td>
<td>.010</td>
<td>.000</td>
</tr>
<tr>
<td>CoSTAR⁸ E</td>
<td>.012</td>
<td>.011</td>
<td>.000</td>
</tr>
<tr>
<td>CoSTAR⁹ L</td>
<td>.021</td>
<td>.009</td>
<td>.008</td>
</tr>
<tr>
<td>CoSTAR⁹ E</td>
<td>.029</td>
<td>.011</td>
<td>.009</td>
</tr>
<tr>
<td>NEURAL L</td>
<td>.039</td>
<td>.046</td>
<td>.044</td>
</tr>
<tr>
<td>NEURAL E</td>
<td>.039</td>
<td>.059</td>
<td>.039</td>
</tr>
<tr>
<td>BIER L</td>
<td>.011</td>
<td>.024</td>
<td>.018</td>
</tr>
<tr>
<td>BIER E</td>
<td>.040</td>
<td>.058</td>
<td>.043</td>
</tr>
</tbody>
</table>

Notes: a. tests based on Section 3; b. tests based on Section 4; c. "L" and "E" denote logistic and exponential tests; d. values denote rejection frequencies at the 5%-level.

### Table 2

<table>
<thead>
<tr>
<th>H₁, n = 100</th>
<th>H₁⁹</th>
<th>H₁⁸</th>
<th>H₁⁷</th>
<th>H₁⁶</th>
<th>H₁⁵</th>
<th>H₁⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td>CoSTAR⁵ L</td>
<td>.236</td>
<td>.401</td>
<td>.383</td>
<td>.247</td>
<td>.486</td>
<td>.194</td>
</tr>
<tr>
<td>CoSTAR⁵ E</td>
<td>.318</td>
<td>.533</td>
<td>.529</td>
<td>.361</td>
<td>.622</td>
<td>.242</td>
</tr>
<tr>
<td>CoSTAR⁷ L</td>
<td>.244</td>
<td>.448</td>
<td>.333</td>
<td>.242</td>
<td>.534</td>
<td>.213</td>
</tr>
<tr>
<td>CoSTAR⁷ E</td>
<td>.394</td>
<td>.695</td>
<td>.562</td>
<td>.431</td>
<td>.813</td>
<td>.302</td>
</tr>
<tr>
<td>CoSTAR⁸ L</td>
<td>.181</td>
<td>.425</td>
<td>.324</td>
<td>.203</td>
<td>.438</td>
<td>.234</td>
</tr>
<tr>
<td>CoSTAR⁸ E</td>
<td>.152</td>
<td>.406</td>
<td>.327</td>
<td>.174</td>
<td>.366</td>
<td>.175</td>
</tr>
<tr>
<td>CoSTAR⁹ L</td>
<td>.114</td>
<td>.295</td>
<td>.232</td>
<td>.198</td>
<td>.214</td>
<td>.096</td>
</tr>
<tr>
<td>CoSTAR⁹ E</td>
<td>.114</td>
<td>.312</td>
<td>.251</td>
<td>.194</td>
<td>.215</td>
<td>.091</td>
</tr>
<tr>
<td>NEURAL L</td>
<td>.158</td>
<td>.372</td>
<td>.302</td>
<td>.085</td>
<td>.028</td>
<td>.119</td>
</tr>
<tr>
<td>NEURAL E</td>
<td>.149</td>
<td>.447</td>
<td>.371</td>
<td>.135</td>
<td>.202</td>
<td>.197</td>
</tr>
<tr>
<td>BIER L</td>
<td>.158</td>
<td>.372</td>
<td>.302</td>
<td>.085</td>
<td>.028</td>
<td>.119</td>
</tr>
<tr>
<td>BIER E</td>
<td>.149</td>
<td>.447</td>
<td>.371</td>
<td>.135</td>
<td>.202</td>
<td>.197</td>
</tr>
<tr>
<td>Table 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>( H_1, n = 500 )</td>
<td>( H_{1L}^R )</td>
<td>( H_{1S}^L )</td>
<td>( H_{1E}^L )</td>
<td>( H_{1BL_1}^R )</td>
<td>( H_{1BL_2}^L )</td>
<td></td>
</tr>
<tr>
<td>CoSTAR((e)) _L</td>
<td>.186</td>
<td>.682</td>
<td>.577</td>
<td>.442</td>
<td>.595</td>
<td>.549</td>
</tr>
<tr>
<td>CoSTAR((e)) _E</td>
<td>.300</td>
<td>.710</td>
<td>.672</td>
<td>.619</td>
<td>.784</td>
<td>.588</td>
</tr>
<tr>
<td>CoSTAR((nd)) _L</td>
<td>.249</td>
<td>.823</td>
<td>.441</td>
<td>.467</td>
<td>.717</td>
<td>.562</td>
</tr>
<tr>
<td>CoSTAR((nd)) _E</td>
<td>.342</td>
<td>.844</td>
<td>.676</td>
<td>.692</td>
<td>.894</td>
<td>.739</td>
</tr>
<tr>
<td>CoSTAR((sc)) _L</td>
<td>.101</td>
<td>.645</td>
<td>.678</td>
<td>.455</td>
<td>.510</td>
<td>.652</td>
</tr>
<tr>
<td>CoSTAR((sc)) _E</td>
<td>.082</td>
<td>.621</td>
<td>.601</td>
<td>.449</td>
<td>.448</td>
<td>.602</td>
</tr>
<tr>
<td>CoSTAR((b)) _L</td>
<td>( \text{NEURAL} _L )</td>
<td>.081</td>
<td>.521</td>
<td>.364</td>
<td>.494</td>
<td>.476</td>
</tr>
<tr>
<td>CoSTAR((b)) _E</td>
<td>( \text{NEURAL} _E )</td>
<td>.111</td>
<td>.466</td>
<td>.364</td>
<td>.507</td>
<td>.486</td>
</tr>
<tr>
<td>( \text{BIER} _L )</td>
<td>.134</td>
<td>.589</td>
<td>.566</td>
<td>.444</td>
<td>.032</td>
<td>.608</td>
</tr>
<tr>
<td>( \text{BIER} _E )</td>
<td>.131</td>
<td>.581</td>
<td>.592</td>
<td>.522</td>
<td>.223</td>
<td>.640</td>
</tr>
</tbody>
</table>

| Table 4 |
|---|---|---|---|---|---|
| \( H_1, n = 1000 \) | \( H_{1L}^R \) | \( H_{1S}^L \) | \( H_{1E}^L \) | \( H_{1BL_1}^R \) | \( H_{1BL_2}^L \) |
| CoSTAR\((e)\) _L | .214 | .739 | .811 | .620 | .638 | .722 |
| CoSTAR\((e)\) _E | .296 | .781 | .820 | .683 | .822 | .748 |
| CoSTAR\((nd)\) _L | .251 | .848 | .626 | .595 | .824 | .892 |
| CoSTAR\((nd)\) _E | .328 | .902 | .864 | .874 | .933 | .929 |
| CoSTAR\((sc)\) _L | .093 | .788 | .892 | .833 | .555 | .738 |
| CoSTAR\((sc)\) _E | .064 | .771 | .825 | .817 | .514 | .711 |
| CoSTAR\((b)\) _L | \( \text{NEURAL} \_L \) | .081 | .538 | .602 | .576 | .584 | .496 |
| CoSTAR\((b)\) _E | \( \text{NEURAL} \_E \) | .081 | .479 | .556 | .566 | .566 | .478 |
| \( \text{BIER} \_L \) | .101 | .720 | .726 | .625 | .042 | .735 |
| \( \text{BIER} \_E \) | .120 | .715 | .784 | .680 | .216 | .741 |
Appendix 3: Formal Proofs

**Proof of Lemma 1.** The proof follows almost directly from Theorem 1 of Bierens and Ploberger (1997): see also Stinchcombe and White (1998). We provide here sufficient details. Theorem 1 of Bierens and Ploberger (1997) states for any $\epsilon$, $E|\epsilon| < \infty$, and bounded vector $x$, if $P(E|\epsilon x| = 0) < 1$ then the set

$$S = \{ \gamma \in \Gamma : E[\epsilon F(\gamma'x)] = 0, P(\gamma'x \in R_0) = 1 \}, i = 1...k$$

is countable and therefore has Lebesgue measure zero and is nowhere dense in $\mathbb{R}^k$, where the real-line interval $R_0$ is defined by Assumption A. By Assumptions 1 and 3, $x$ and $h(x,\delta)$ are $\mathfrak{F}$-measurable and bounded, $0 < \inf_{\delta \in \Delta} |h(x,\delta)| \leq \sup_{\delta \in \Delta} |h(x,\delta)| < \infty$ with probability one. Thus, if

$$P(E[(y - f(x,\theta_0))]|x| = 0) = 1$$

for some $\theta_0 \in \Theta$ such that the null is true, then

$$P(E[h(x,\delta)|x| = 0) = P(h(x,\delta)E[|\epsilon|x] = 0) = P(E[|\epsilon]|x = 0) = 1.$$  \hspace{1cm} (46)

Under the alternative, we likewise deduce

$$P(E[h(x,\delta)|x| = 0) = P(h(x,\delta)E[|\epsilon|x] = 0) = P(E[|\epsilon]|x = 0) < 1.$$  \hspace{1cm} (47)

Moreover, by assumption $h(x,\delta)F(\gamma'x)$ satisfies Assumption A. Hence, in tandem with (48) and (49), we may simply re-define $\epsilon$ as $h(x,\delta)\epsilon$ in (46) and apply Theorem 1 of Bierens and Ploberger (1997): each set $S_i$ in (9) is countable and therefore has Lebesgue measure zero. \hfill \blacksquare

**Proof of Theorem 2.** Consider (i), and write $F_i = F_i(\gamma) = F(\gamma'x_i)$. Applying an exact Taylor expansion and invoking the mean-value theorem (see, e.g., Amemiya, 1985), there exist vectors $\theta \in (\hat{\theta},\theta_0)$ and $\theta^* \in (\hat{\theta},\theta_0)$ such that
the moment condition can be written as

\begin{equation}
\dot{s}(\gamma, \delta)
= \frac{1}{n} \sum_{t=1}^{n} (y_t - f(x_t, \hat{\theta})) \dot{F}_t
= \frac{1}{n} \sum_{t=1}^{n} \dot{F}_t(y_t - f(x_t, \hat{\theta}))
= \frac{1}{n} \sum_{t=1}^{n} \dot{F}_t \left( y_t - f(x_t, \theta_0) - \frac{\partial}{\partial \theta} f(x_t, \theta) |_{\theta = \hat{\theta}} \left( \hat{\theta} - \theta_0 \right) \right)
= \frac{1}{n} \sum_{t=1}^{n} \dot{F}_t \left( \epsilon_t - \frac{\partial}{\partial \theta} f(x_t, \theta_0) |_{\theta = \hat{\theta}} \dot{A}^{-1} |_{\theta = \theta_0} \frac{1}{n} \sum_{s=1}^{n} \epsilon_s \frac{\partial}{\partial \theta} f(x_s, \theta_0) \right)
= \frac{1}{n} \sum_{t=1}^{n} \dot{F}_t \epsilon_t - \frac{1}{n} \sum_{t=1}^{n} \dot{F}_t \frac{\partial}{\partial \theta} f(x_t, \theta) |_{\theta = \hat{\theta}} \dot{A}^{-1} |_{\theta = \theta_0} \frac{1}{n} \sum_{t=1}^{n} \epsilon_t \frac{\partial}{\partial \theta} f(x_t, \theta_0)
= \frac{1}{n} \sum_{t=1}^{n} \left( \dot{F}_t - \beta_{\theta = \hat{\theta}} \dot{A}^{-1} |_{\theta = \theta_0} \frac{\partial}{\partial \theta} f(x_t, \theta_0) \right) \epsilon_t
= \frac{1}{n} \sum_{t=1}^{n} \dot{g}_t(\theta, \gamma, \delta) |_{\theta_0, \theta} \epsilon_t
\end{equation}

say, where

\begin{align}
\dot{A} &= \dot{A}(\theta) = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} f(x_t, \theta) \frac{\partial}{\partial \theta} f(x_t, \theta) \\
&\quad - \frac{1}{n} \sum_{t=1}^{n} \epsilon_t \frac{\partial^2}{\partial \theta \partial \theta} f(x_t, \theta) \\
\dot{b} &= \dot{b}(\theta, \gamma, \delta) = \frac{1}{n} \sum_{t=1}^{n} \dot{F}_t(\gamma, \delta) \frac{\partial}{\partial \theta} f(x_t, \theta) \\
\dot{g}_t(\theta, \gamma, \delta) |_{\theta_0, \theta} &\equiv \dot{F}_t - \dot{b}(\theta, \gamma, \delta) |_{\theta = \hat{\theta}} \dot{A}(\theta)^{-1} |_{\theta = \theta_0} \frac{\partial}{\partial \theta} f(x_t, \theta_0).
\end{align}

By Assumption 1, \( \epsilon_t \) is a \((0, \sigma^2)\)-martingale difference sequence under the null. Because \( x_t \) and \( \dot{F}_t \) are \( \tilde{g}_{t-1} \)-measurable it follows that \( \dot{g}_t(\theta, \gamma, \delta) \) is \( \tilde{g}_{t-1} \)-measurable, hence \( \dot{g}_t(\theta, \gamma, \delta) \epsilon_t \) forms a vector martingale difference sequence for any \( \theta, \gamma, \delta \):

\begin{equation}
E(\dot{g}_{t,i}(\theta, \gamma, \delta) \epsilon_t | \tilde{g}_{t-1}) = 0, \ i = 1 \ldots k.
\end{equation}

Therefore, by Assumptions 1-4, the Slutsky Theorems and the martingale central limit theorem, cf. McLeish (1974), the sequence \( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \dot{g}_t(\theta, \gamma, \delta) |_{\theta_0, \theta} \epsilon_t \) converges in law jointly to a Gaussian random vector. In particular,

\begin{equation}
\sqrt{n} \dot{s}(\gamma, \delta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \dot{g}_t(\theta, \gamma, \delta) |_{\theta_0, \theta} \epsilon_t \quad \Longrightarrow \quad N(0, V(\gamma, \delta)),
\end{equation}

pointwise in \( \Gamma \) and \( \Delta \) for some covariance matrix, \( V(\gamma, \delta) \).
The covariance matrix \( V(\gamma, \delta) \) will be the point-wise probability limit of

\[
n\hat{s}(\gamma, \delta)\hat{s}(\gamma, \delta)'
\]

\[
= \left[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \hat{g}_t(\theta, \gamma, \delta) \right] \left[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \hat{g}_t(\theta, \gamma, \delta) \right]'
\]

provided the limit exists. In particular, by the martingale difference property of the innovations \( \epsilon_t \) and \( \hat{g}_t(\theta, \gamma, \delta)\epsilon_t \) under \( H_0 \), we deduce

\[
= \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2 \hat{g}_t(\theta, \gamma, \delta) |_{\hat{\theta}, \theta^*} \hat{g}_t(\theta, \gamma, \delta) |_{\hat{\theta}, \theta^*} + O_p(1)
\]

\[
= \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2 \left( \hat{F}_t - \hat{b}_{|\theta=\hat{\theta}} \hat{A}^{-1} |_{\theta=\theta^*} \frac{\partial}{\partial \theta} f(x_t, \theta_0) \right)
\]

\[
\times \left( \hat{F}_t - \hat{b}_{|\theta=\hat{\theta}} \hat{A}^{-1} |_{\theta=\theta^*} \frac{\partial}{\partial \theta} f(x_t, \theta_0) \right)' + O_p(1),
\]

where the term \( O_p(1) \) contains the cross-products \( \epsilon_s \epsilon_t, s \neq t \), and follows from ergodicity and the \((0, \sigma^2)\)-martingale-difference property of \( \epsilon_t \). By stationarity, Assumptions 1 - 4, and standard properties of functional probability limits we deduce by a weak law of large numbers (see, e.g., Jenrich, 1969; White, 1982)

\[
\left[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \hat{g}_t(\theta, \gamma, \delta) \right] \left[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \hat{g}_t(\theta, \gamma, \delta) \right]'
\]

\[
\rightarrow E\epsilon_t^2 \left\{ F_t(\gamma) h_t(\delta) - b(\gamma, \delta) A^{-1} \frac{\partial}{\partial \theta} f(x_t, \theta_0) \right\}
\]

\[
\times \left\{ F_t(\gamma) h_t(\delta) - b(\gamma, \delta) A^{-1} \frac{\partial}{\partial \theta} f(x_t, \theta_0) \right\}'.
\]

Finally, consider \((ii)\). Under \( H_1 \) the innovations \( \epsilon_t \) form a white noise process, hence \( \hat{\theta} \) is a consistent estimator of the \( k \)-vector \( \theta \). Thus, (13) follows from Lemma 1 and the strong law of large numbers. In particular, under \( H_1 \),

\[
\frac{1}{n} \sum_{t=1}^{n} (y_t - f(x_t, \hat{\theta})) h_{i,t}(\delta) F_t(\gamma) \rightarrow E(\epsilon_t h_{i,t}(\delta) F_t(\gamma)) \neq 0
\]

for each \( i = 1...k \) with probability one. Therefore, \( \eta_i(\gamma) = E(\epsilon_t h_{i,t}(\delta) F_t(\gamma)) \neq 0 \), a.s. \( \blacksquare \)

**Proof of Lemma 3.** The result holds if \( \hat{\Psi}(x) \) is any bounded, measurable one-to-one function, cf. Remark 1 of Lemma 1. Assume \( \hat{\Psi}(x) \) is not one-to-one. Then if \( F(\gamma'\hat{\Psi}(x)) = h(x, \delta) F(\gamma'\Psi(x)) \) for some \( F(\gamma'\Psi(x)) \) that satisfies Assumption A and some \( h(x, \delta) \) that is bounded and non-zero with probability one, then by Lemma 1 the set

\[
S = \{ \gamma \in \Gamma : E[h(x, \delta) F(\gamma'\Psi(x))] = 0, \text{ and } P(\gamma' x \in R_0) = 1 \}, i = 1...k,
\]

27
has Lebesgue measure zero and is nowhere dense. □

**Proof of Theorem 4.** The claim follows from the following three lemmas.

Recall we define \( x_t(\delta) \equiv (\delta_1, x_{t,1}, ..., \delta_k, x_{t,k})' \) and \( \Psi(x_t(\delta)) \equiv (\exp(x_{t,1}(\delta)), ..., \exp(x_{t,k}(\delta)))' \) for \( \delta_i \neq 0, i = 1..k \), and

\[
S_\delta = \{ \gamma \in \Gamma : E[\epsilon_t F(\gamma'\Psi(x_t(\delta)))] = 0 \text{ and } P(\gamma'\Psi(x_t(\delta)) \in R_0) = 1 \}.
\]

For brevity, assume \( x_t \) does not contain a constant term.

**Lemma A1.** Let \( \epsilon_t \) be a random variable satisfying \( E[|\epsilon_t|] < \infty \), and let \( x_t \) be an \( \mathfrak{s}_t \)-measurable bounded vector in \( \mathbb{R}^k \), \( 0 < k < \infty \), such that \( P[|E(\epsilon_t|x_t)| = 0] < 1 \). Assume \( \Gamma \) is a bounded subset of \( \mathbb{R}^k \), and let Assumption A' hold for \( F(u) \). Then the set \( S_\delta \) has Lebesgue measure and is nowhere dense in \( \mathbb{R}^k \) uniformly in \( \Delta \), \( \delta_i \neq 0, i = 1..k \).

**Lemma A2.** Let the assumptions of Lemma A1 hold. If \( P[E(\epsilon_t|x_t)| = 0] < 1 \), then for each point \( (\gamma, \delta) \in \Gamma \times \Delta \), \( \delta_i \neq 0, i = 1..k \), and any integer \( k \)-vector of integers \( r \in \mathbb{Z}^k \) there exists a vector \( m \in \mathbb{Z}^k \), \( m \geq r \), such that

\[
E \left[ \epsilon_t \exp\{m'x_t(\delta)\} \prod_{i=1}^k F^{|m_i-r_i|}(u) | u = \gamma'\Psi(x_t(\delta)) \right] \neq 0. \tag{58}
\]

**Lemma A3.** If \( P[E(\epsilon_t|x_t)| = 0] < 1 \), then for every vector \( m_0 \in \mathbb{Z}^k \), \( m_0 \geq r \), such that Lemma A2 holds for \( m = m_0 \), there exists another vector \( m_1 \) such that Lemma A2 holds for \( m = m_1 \).

By Lemma A2, for every point \( (\gamma, \delta) \in \Gamma \times \Delta \), \( \delta_i \neq 0, i = 1..k \), there exists at least one vector \( m \in \mathbb{Z}^k \) such that the moment condition

\[
E \left[ \epsilon_t \exp\{m'x_t(\delta)\} \prod_{i=1}^k F^{|m_i-r_i|}(u) | u = \gamma'\Psi(x_t(\delta)) \right] \neq 0 \tag{59}
\]

holds. Invoking Lemma A3, there must therefore exist infinitely many such vectors \( m \). □

**Proof of Lemma A1.** The claim follows directly from Lemma 3: for any bounded \( \delta_i \neq 0, i = 1..k \), the mapping \( \Psi(x_t(\delta)) \equiv (\exp(x_{t,1}(\delta)), ..., \exp(x_{t,k}(\delta)))' \) is one-to-one and therefore generates the same \( \sigma \)-algebra as \( x_t \), hence Lemma 1 holds for \( \Psi(x_t(\delta)) \). □

**Proof of Lemma A2.** Let \( H_1 \) hold, and consider \( k = 1 \). For any point \( (\gamma, \delta_0), \delta_0 \neq 0 \), and any integer \( r \in \mathbb{Z} \), expand \( E[\epsilon_t \exp\{r\delta_0 x_t + \gamma e^{\delta_0 x_t}\}] \) around an arbitrary point \( \gamma_0 \):

\[
E[\epsilon_t \exp\{r\delta_0 x_t\} F(\gamma'\Psi(x_t(\delta_0)))]
\]

\[
= \sum_{i=0}^\infty \left( \frac{\partial}{\partial \gamma} \right)^i E[\epsilon_t \exp\{r\delta_0 x_t\} F(\gamma'\Psi(x_t(\delta_0)))] \bigg|_{\gamma = \gamma_0} \frac{(\gamma - \gamma_0)^i}{i!}
\]

\[
= \sum_{i=0}^\infty E[\epsilon_t \exp\{(r+i)\delta_0 x_t\} F^i(\gamma_0'\Psi(x_t(\delta_0)))] \frac{(\gamma - \gamma_0)^i}{i!},
\]

\[
\text{E}[\epsilon_t \exp\{r\delta_0 x_t\} F(\gamma'\Psi(x_t(\delta_0)))]
\]
where \( F^i(u) = (\partial/\partial u)^i F(u) \). For \( m \geq r \), if every \((m - r)^{th}\) partial derivative at the point \((\gamma_0, \delta_0)\) is identically zero,

\[
\left( \frac{\partial}{\partial \gamma} \right)^{m-r} E [\epsilon_t \exp \{r \delta_0 x_t\} F (\gamma' \Psi(x_t(\delta_0)))] \big|_{\gamma=\gamma_0} \quad (61)
\]

then, by (61), at the point \((\gamma, \delta_0)\) for any \( \gamma \) the moment condition \( E [\epsilon_t \exp \{r \delta_0 x_t\} F (\gamma' \Psi(x_t(\delta_0)))] \) reduces to zero:

\[
E [\epsilon_t \exp \{r \delta_0 x_t\} F (\gamma' \Psi(x_t(\delta_0)))] = 0, \quad (62)
\]

Notice that

\[
P (E[\epsilon_t \exp \{r \delta_0 x_t\} | x_t] = 0) = P (E[\epsilon_t | x_t] = 0) < 1, \quad (63)
\]

hence Lemma A1 applies to \( \epsilon_t \exp \{r \delta_0 x_t\} \); if for any \( \delta_0 \neq 0 \) the moment condition \( E[\epsilon_t \exp \{r \delta_0 x_t\} F (\gamma' \Psi(x_t(\delta_0)))] = 0 \) for some \( \gamma \), then \( E [\epsilon_t \exp \{r \delta_0 x_t\} F (\gamma' \Psi(x_t(\delta_0)))] \neq 0 \) for \( \gamma \) arbitrarily close to \( \gamma_0 \), hence (63) cannot hold for all \( \gamma \). Therefore, for every point \((\gamma_0, \delta_0)\) and any \( r \in \mathbb{Z} \) there exists at least one integer \( m \geq r \) such that

\[
\left( \frac{\partial}{\partial \gamma} \right)^{m-r} E [\epsilon_t \exp \{r \delta_0 x_t\} F (\gamma' \Psi(x_t(\delta_0)))] \big|_{\gamma=\gamma_0, \delta_0} \quad (64)
\]

In the multivariate case, \( k \geq 1 \), an essentially identical argument applies. If for some point \((\gamma_0, \delta_0), \delta_{0,i} \neq 0, i = 1...k, \) and \( k \)-vector \( r \in \mathbb{Z}^k \), and every integer \( k \)-vector \( m \geq r \)

\[
\Pi_{i=1}^k \left( \frac{\partial}{\partial \gamma_i} \right)^{m_i-r_i} E \left[ \epsilon_t \exp \left\{ \sum_{j=1}^k r_j \delta_{0,j} x_{t,j} \right\} F^i (\gamma' \Psi(x_t(\delta_0))) \right] \big|_{\gamma=\gamma_0} \quad (65)
\]

\[
= E \left[ \epsilon_t \exp \left\{ \sum_{j=1}^k m_j \delta_{0,j} x_{t,j} \right\} F^i (\gamma' \Psi(x_t(\delta_0))) \right] = 0,
\]

then, from an infinite order Taylor expansion we deduce for any \( r \in \mathbb{Z}^k \) and every \( k \)-vector \( \gamma \in \Gamma \)

\[
E \left[ \epsilon_t \exp \left\{ \sum_{i=1}^k r_i \delta_{0,i} x_{t,i} \right\} F (\gamma' \Psi(x_t(\delta_0))) \right] = 0, \quad (66)
\]

again, a contradiction of Lemma A1. Therefore, for each point \((\gamma_0, \delta_0) \in \Gamma \times \Delta, \delta_{0,i} \neq 0, \) and integer vector \( r \in \mathbb{Z}^k \) there exists at least one vector \( m \geq r \) of
integers such that

\[ \prod_{i=1}^{k} \left( \frac{\partial}{\partial \gamma_i} \right)^{m_i-r_i} E \left[ \epsilon_i \exp \left\{ \sum_{j=1}^{k} r_j \delta_{0,j} x_{t,j} \right\} F' \left( \gamma' \Psi(x_t(\delta_0)) \right) \right] \mid \gamma=\gamma_0 \]  

(67)

\[ = E \left[ \epsilon_i \exp \left\{ \sum_{j=1}^{k} m_j \delta_{0,j} x_{t,j} \right\} \prod_{i=1}^{k} F^{m_i-r_i}(u) \mid u=\gamma_0 \Psi(x_t(\delta_0)) \right] \neq 0, \]

which completes the proof. ■

**Proof of Lemma A3.** Let \( H_1 \) hold, and let \( k = 1 \). For any point \((\gamma_0, \delta_0) \in \Gamma \times \Delta, \delta_0 \neq 0, \) and any integer \( r \in \mathbb{Z} \) assume Lemma A2 holds:

\[ E \left[ \epsilon_i \exp \{ m \delta_0 x_t \} F^{m-r} (\gamma_0' \Psi(x_t(\delta_0))) \right] \neq 0 \]  

(68)

for some \( m = m_{j-1} \), where \( m_{j-1} \in \mathbb{Z}, m_{j-1} \geq r \). We prove that there must exist a larger integer \( m_j > m_{j-1} \) such that (69) holds for \( m = m_j \).

Now, for **some** integer \( \tilde{m} > m_{j-1} \) and **some** point \( \gamma \in \Gamma \) the moment condition

\[ E[\epsilon_i \exp \{ \tilde{m} \delta_0 x_t \} F^{\tilde{m}-r} (\gamma' \Psi(x_t(\delta_0)))] \neq 0 \]  

(69)

must hold: by Assumption A' infinitely many \( F^s(u) \) are non-zero with probability one and satisfy Assumption A', therefore for some \( \tilde{m} \) if

\[ E[\epsilon_i \exp \{ \tilde{m} \delta_0 x_t \} F^{\tilde{m}-r} (\gamma' \Psi(x_t(\delta_0)))] = 0 \]  

(70)

for any point \( \tilde{\gamma} \), then

\[ E[\epsilon_i \exp \{ \tilde{m} \delta_0 x_t \} F^{\tilde{m}-r} (\gamma' \Psi(x_t(\delta_0)))] = 0 \]  

(71)

for infinitely many \( \gamma \) is a dense neighborhood of \( \tilde{\gamma} \). Thus, (70) holds for some pair \( \tilde{m} \) and \( \gamma \).

Expand the moment condition \( E[\epsilon_i \exp \{ \tilde{m} \delta_0 x_t \} F^{\tilde{m}-r} (\gamma' \Psi(x_t(\delta_0)))] \) around \( \gamma_0 \):

\[ E \left[ \epsilon_i \exp \{ \tilde{m} \delta_0 x_t \} F (\gamma' \Psi(x_t(\delta_0))) \right] \]  

(72)

\[ = \sum_{i=0}^{\infty} E \left[ \epsilon_i \exp \{ \tilde{m} \delta_0 x_t \} F^{\tilde{m}+i-r} (\gamma_0' \Psi(x_t(\delta_0))) (\gamma - \gamma_0)^i / i! \right] \neq 0. \]

Because (70) holds for some \( \gamma \) and some \( \tilde{m} > m_{j-1} \geq r \), we deduce from (73) there must exist another integer \( m_j \geq \tilde{m} > m_{j-1} \) such that

\[ E \left[ \epsilon_i \exp \{ m_j \delta_0 x_t \} F^{m_j-r} (\gamma_0' \Psi(x_t(\delta_0))) \right] \neq 0, \]

(73)

at \( m = m_j \). This proves the result for \( k = 1 \). The multivariate case, \( k \geq 1 \), follows similarly. ■

**Proof of Theorem 6.** The result follows immediately from Theorem 7.

Using notation from that result, define

\[ g(x_t, \gamma, \delta, m, r) = I_k \times \prod_{i=1}^{k} F^{m_i-r_i} (u) \mid u=\gamma \Psi(x_t(\delta)) \]  

(74)

\[ h(x_t, \gamma, \delta, m, r) = g(x_t, \gamma, \delta, m, r) \times \partial / \partial \theta f (x_t, \theta_0). \]

30
Then moment condition (22) can be represented as
\[
E \left[ \epsilon_t \frac{\partial}{\partial \theta} f(x_t, \theta_0) \exp \{ m' x_t(\delta) \} \prod_{i=1}^{k} F^{m_i-r_i}(\gamma' \Psi(x_t(\delta))) \right] \tag{75}
\]
\[
= E \left[ \epsilon_t h(x_t, \gamma, \delta, m, r) \exp \{ \delta' x_t(m) \} \right].
\]
In particular, if \( m \neq 0 \), then for any \( q \in \mathbb{R}^k, q \neq 0, \)
\[
q' g(x_t, \gamma, \delta, m, r) = q' I_k \times \prod_{i=1}^{k} F^{m_i-r_i}(u) \big|_{u=\gamma' \Psi(x_t(\delta))} \tag{76}
\]
never has a degenerate distribution for any \( \delta_i \neq 0, i = 1 \ldots k \), hence Theorem 7.ii applies.
If \( m = 0 \), then simply re-define
\[
g(x_t, \delta, m) = I_k \times \exp \{ m' x_t(\delta) \} \tag{77}
\]
\[
h(x_t, \delta, m) = g(x_t, \delta, m) \times \frac{\partial}{\partial \theta} f(x_t, \theta_0),
\]
hence moment condition (22) can be represented as
\[
E \left[ \epsilon_t h(x_t, \delta, m, r) \tilde{F}(\gamma' \Psi(x_t(\delta), r, m) \right] \tag{78}
\]
\[
= \tilde{F}(u, r, m) = \prod_{i=1}^{k} F^{m_i-r_i}(u).
\]
By Assumption A’, because infinitely many \( F^*(u) \neq 0 \) with probability one, \( \tilde{F}(u, r, m) \neq 0 \) with probability one for infinitely many \( m \) and \( r \) and \( \tilde{F}(u, r, m) \) satisfies Assumption A’. Moreover, \( g(x_t, \delta, m) = I_k \times \exp \{ 0' x_t(\delta) \} = I_k \) is constant-valued and trivially has degenerate marginal distributions, hence Theorem 7.iii applies. \( \blacksquare \)

**Proof of Corollary 7.** Most of the steps of the present proof mimic the line of proof of Corollary 1 of Bierens (1990). We present the details here for clarity. Write \( \tilde{F}_{t,j} = \tilde{F}(\delta, \gamma, m) = h(x_t, \delta, j)F(\gamma' \Psi(x_t, m)) \). For \( n = 1, 2, \ldots, \)
define the scalar
\[
f_n(x_t) = \alpha_{n,0} + \sum_{j=1}^{\infty} \alpha'_{n, j} \tilde{F}_{t,j}, \tag{79}
\]
where \( \alpha_{n,0} \) is a scalar, and the \( m_2 \)-vectors \( \alpha_{n, j} \) satisfy
\[
E[f_n(x_t)] = 0, \quad E \left[ f_n(x_t) \tilde{F}_{t,j} \right] = 0, \quad j < n. \tag{80}
\]
Define for \( n > 0 \)
\[
\psi_n(x_t) = f_n(x_t)/(E[f_n(x)]^2)^{1/2}, \quad \text{if } E[f_n(x_t)]^2 > 0 \tag{81}
\]
\[
\psi_n(x_t) = 0, \quad \text{if } E[f_n(x_t)]^2 = 0.
\]
By construction, the nonzero functions \( \psi_n(x_t) \) form an orthonormal Hilbert space in \( L_2(\Omega, \mathcal{F}_t, P) \). Derive the \( L_2 \)-metric projection of \( y_t \) onto the closed
linear span, \( \overline{\mathcal{F}}(\psi_n(x_t)) \equiv \overline{\mathcal{F}}(\psi_n(x_t)); \ n \geq 0 \). In particular, there exists a sequence of functions \( \gamma_n \) such that

\[
P(y_t|\overline{\mathcal{F}}(\psi_n(x_t))) = \sum_{n=0}^{\infty} \gamma_n \psi_n(x_t),
\]

where the infinite series is well defined for the orthonormal Hilbert space defined by \( \psi_n(x_t) \); see Royden (1968: Proposition 27). By the \( L_2 \)-metric projection problem, we deduce \( \gamma_n \) satisfies

\[
E \left[ y_t - \sum_{n=0}^{\infty} \gamma_n \psi_n(x_t) \right] \psi_j(x_t) = 0, \ j = 0, 1, \ldots,
\]

and because for \( j < n \), cf. (80) and (81)

\[
E \left[ \psi_n(x_t) \psi_j(x_t) \right] = \frac{E \left[ f_n(x_t) f_j(x_t) \right]}{\left( E \left[ f_n(x) \right]^2 \right)^{1/2} \left( E \left[ f_j(x) \right]^2 \right)^{1/2}} = 0,
\]

we deduce recursively \( \gamma_n = E[y_t \psi_n(x_t)] \). From (81), (82) and (84), we may write \( P(y_t|\overline{\mathcal{F}}(\psi_n(x_t))) \) as

\[
P(y_t|\overline{\mathcal{F}}(\psi_n(x_t)))
= \sum_{n=0}^{\infty} \gamma_n \psi_n(x_t) = \sum_{n=0}^{\infty} \left[ \gamma_n / \left( E[f_n(x)^2] \right)^{1/2} \right] f_n(x_t)
= \sum_{n=0}^{\infty} \left[ \gamma_n / \left( E[f_n(x)^2] \right)^{1/2} \right] \left( \alpha_n, 0 + \sum_{j=1}^{n} \alpha_{n,j} \tilde{F}_{t,j} \right)
= \beta_{0,0} + \sum_{n=1}^{\infty} \left[ \beta_{n,0} + \sum_{j=1}^{n} \beta'_{n,j} \tilde{F}_{t,j} \right].
\]

By orthogonality of the metric projection error \( y_t - P(y_t|\overline{\mathcal{F}}(\psi_n(x_t))) \), and orthonormality of \( \psi_n(x_t) \), we deduce

\[
E \left[ y_t - P(y_t|\overline{\mathcal{F}}(\psi_n(x_t))) \right] \tilde{F}_t(\delta_j, \gamma_j, m_j) = 0, \ j = 1, 2, \ldots
\]

for any sequences of \( \delta_j, \gamma_j \neq 0 \), and \( m_j \). Because equality holds for every sequence of \( k \)-vector integers \( m_j \), we deduce by Corollary 5 that \( E \left[ y_t - P(y_t|\overline{\mathcal{F}}(\psi_n(x_t))) \right] | x_t \) = 0 with probability one, hence

\[
E \left[ y_t | x_t \right] = P(y_t|\overline{\mathcal{F}}(\psi_n(x_t)))
= \beta_{0,0} + \sum_{n=1}^{\infty} \left[ \beta_{n,0} + \sum_{j=1}^{n} \beta'_{n,j} \tilde{F}_{t,j} \right],
\]

which completes the proof. \( \blacksquare \)

**Proof of Theorem 8.** We assume \( x_t \) does not contain a constant term. Recall

\[
V(\gamma, \delta) = E \left[ z_t(\gamma, \delta) z_t(\gamma, \delta)' e_t^2 \right]
\]

where we define the \( k \)-vector \( z_t(\gamma, \delta) \) as

\[
z_t(\gamma, \delta) = \tilde{F}_t(\gamma, \delta) - b(\gamma, \delta) A^{-1} \partial \theta f(x_t, \theta_0).
\]
For simplicity, we will drop the arguments $\gamma$ and $\delta$ from $z_t(\gamma, \delta)$.

Using (27), $h(x_t, \delta) = g_t \times \partial/\partial \theta f(x_t, \theta_0)$, where $g_t = g(x_t, \delta)$ for some $m_2 \times k$ matrix $g(x_t, \delta)$, we write

$$z_t = F_t - b A^{-1} \frac{\partial}{\partial \theta} f(x_t, \theta_0)$$

$$= F_t h_t - b A^{-1} \frac{\partial}{\partial \theta} f(x_t, \theta_0)$$

$$= F_t g_t - b A^{-1} \frac{\partial}{\partial \theta} f(x_t, \theta_0)$$

$$= (F_t g_t - b A^{-1}) \partial/\partial \theta f(x_t, \theta_0).$$

**Step 1** ($r'V(\gamma, \delta)r = 0$): Any choice of $\gamma \in S^*$ implies there exists a vector $r \in \mathbb{R}^{m_2}$, $r \neq 0$ such that\footnote{Note that $r'V(\gamma, \delta)r = r'E \left[z_t(\gamma, \delta) z_t(\gamma, \delta)' E \left(\epsilon_t^2 | x_t \right) \right] \geq 0$, therefore is suffices to consider only the indefinite case with equality.} $r'V(\gamma, \delta)r = 0$, hence $V(\gamma, \delta)$ is uniformly indefinite. Consider any $\delta \in \Delta$. Then the equality $r'V(\gamma, \delta)r = 0$ reduces to

$$r'V(\gamma, \delta)r = r' E \left[ z_t z_t' E \left(\epsilon_t^2 | x_t \right) \right] r$$

$$= E \left[ r' z_t z_t' E \left(\epsilon_t^2 | x_t \right) \right]$$

$$= E \left[ \left( \sum_{i=1}^{m_2} r_i z_{t,i} \right)^2 E \left(\epsilon_t^2 | x_t \right) \right] = 0.$$

By Assumption C, $E \left(\epsilon_t^2 | x_t \right) > 0$ with probability one, hence the equality holds if and only if for $r \neq 0$ with probability one

$$\sum_{i=1}^{k} r_i z_{t,i} = 0.$$  

**Step 2** ($\gamma = 0$): Separating $z_t$, cf. (90), we obtain from (92) for some $r \neq 0$

$$F_t \sum_{i=1}^{k} \partial/\partial \theta_i f(x_t, \theta_0) \sum_{j=1}^{m_2} [g_t]_{j,i} r_j$$

$$= \sum_{i=1}^{k} \partial/\partial \theta_i f(x_t, \theta_0) \sum_{j=1}^{m_2} [b A^{-1}]_{j,i} r_j$$

with probability one.

We consider three cases: (i) $g_t = g$, a non-zero constant matrix; (ii) $r' g_t$ and $F_t g_{t,i,j}$ have non-degenerate marginal distributions for all $r \neq 0$ and at least one pair $i,j$; (iii) $r' g_t$ has degenerate marginal distributions for some $r \neq 0$, and $F_t g_{t,i,j}$ have non-degenerate marginal distributions for at least one pair $i,j$.\footnote{Note that $r'V(\gamma, \delta)r \geq 0$; therefore is suffices to consider only the indefinite case with equality.}
The origin above. Furthermore, because \( g(x, \delta) = g \) is constant-valued for any \( \delta \), we remove the argument \( \delta \) in the following discourse.

Using \( g_t = g \), we have

\[
F_t \sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} f(x_t, \theta_0) \sum_{j=1}^{m_2} g_{j,i} r_j
\]

or

\[
\sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} f(x_t, \theta_0) [F_t \alpha_i(r) - \beta_i(r, \gamma)] = 0
\]

with probability one for some \( r \neq 0 \), where we define

\[
\alpha_i(r) = \sum_{j=1}^{m_2} g_{j,i} r_j, \quad \beta_i(r, \gamma) = \sum_{j=1}^{m_2} [b(\gamma) A^{-1}]_{j,i} r_j.
\]

The vector \( \beta(r, \gamma) \) uniformly contains at least one non-zero component due to the positive definiteness of \( A = E[(\partial/\partial \theta) f(x_t, \theta)](\partial/\partial \theta') f(x_t, \theta) \), and the fact that the matrix \( b(\gamma) \) has full row and column rank uniformly in \( \Gamma \) by Assumption D. In order to see this, consider the contrary and let \( \beta(r, \gamma) = r' b(\gamma) A^{-1} = 0 \) for some \( r \neq 0 \) and \( \gamma \), hence \( r' b(\gamma) = 0 \), which implies either \( b(\gamma) \) does not have full row rank uniformly in \( \Gamma \), a contradiction of Assumption D; or \( r = 0 \), which is ruled out by construction. Therefore for each \( \gamma \), \( \beta_i(r, \gamma) \neq 0 \) for at least one \( i = 1...k \).

The fact that \( \beta(r, \gamma) \neq 0 \) uniformly in \( \Gamma \) implies

\[
\sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} f(x_t, \theta_0) \beta_i(r, \gamma) \neq 0
\]

with probability one by Assumption D; equality holds \( \sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} f(x_t, \theta_0) \beta_i(r, \gamma) = 0 \) with strictly positive probability if and only if \( \beta(r, \gamma) = 0 \), which is ruled out above. Furthermore, because \( \sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} f(x_t, \theta_0) \beta_i(r, \gamma) \neq 0 \) with probability one, we deduce from equality (94) that \( F_t \sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} f(x_t, \theta_0) \alpha_i(r) \neq 0 \) with probability one, therefore \( \sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} f(x_t, \theta_0) \alpha_i(r) \neq 0 \) with probability one \((F_t \neq 0 \text{ with probability one by Assumption 3})\), hence \( \alpha(r) \neq 0 \) by Assumption D.

Thus, by Assumption D, from (95) the orthogonality

\[
\langle F_t \alpha(r) - \beta(r, \gamma), \partial/\partial \theta f(x_t, \theta_0) \rangle
\]

or

\[
F_t \sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} f(x_t, \theta_0) [F_t \alpha_i(r) - \beta_i(r, \gamma)] = 0
\]

holds if and only if \( F_t \alpha(r) - \beta(r, \gamma) = 0 \) with probability one uniformly in \( \Gamma \). The origin \( \gamma = 0 \) solves equality (98) because

\[
F_t(0) \alpha(r) = c \alpha(r) = crg
\]
for some scalar-constant $c$, cf. Assumption 3, and

\begin{equation}
\beta(r, 0) = r' b(0) A^{-1}
= r' E[F_t(0) g \times \partial/\partial \theta f(x_t, \theta_0) \partial/\partial \theta' f(x_t, \theta_0)] A^{-1}
= r' E[cg \times \partial/\partial \theta f(x_t, \theta_0) \partial/\partial \theta' f(x_t, \theta_0)] A^{-1}
= r' cg \times E[\partial/\partial \theta f(x_t, \theta_0) \partial/\partial \theta' f(x_t, \theta_0)] A^{-1}
= r' cg.
\end{equation}

Finally, because $\alpha(r) \neq 0$ and $\beta(r, \gamma) \neq 0$ uniformly in $\Gamma$, there exists at least one $i$ and $j$ such that $\alpha_i(r) \neq 0$ and $\beta_j(r, \gamma) \neq 0$. Clearly $i = j$: if $i \neq j$ for every $\alpha_i(r) \neq 0$ and $\beta_j(r, \gamma) \neq 0$, then with probability one

\[ F_t \alpha_i(r) = 0 \Rightarrow \alpha_i(r) = 0, \tag{101} \]

because $F_t \neq 0$ with probability one, cf. Assumption 3, which contradicts the assumption $\alpha_i(r) \neq 0$. If $i = j$ for some pair $i, j$ such that $\alpha_i(r) \neq 0$ and $\beta_j(r, \gamma) \neq 0$ uniformly in $\Gamma$, then

\[ F_t = \beta_j(r, \gamma)/\alpha_i(r), \tag{102} \]

a constant. By Assumption 3, $F_t(\gamma)$ is constant-valued if and only if $\gamma = 0$. We conclude in this case $\gamma = 0$ is the only vector that solves $r' V(\gamma, \delta) r = 0$ for $r \neq 0$, $S^* = \{0\}$.

**Case ii:** $[r' g(x_t, \delta)]_i$ and $F_t g_{t,i,j}$ have non-degenerate marginal distributions for at least one $l$ and one pair $i, j$.

We have

\[ \sum_{i=1}^k \partial/\partial \theta_i f(x_t, \theta_0) [F_t \alpha_i(r) - \beta_i(r, \gamma, \delta)] = 0 \tag{103} \]

with probability one for some $r \neq 0$, where

\[ \alpha_i(r, \delta) = \sum_{j=1}^{m_2} g_{t,i,j} r_j, \quad \beta_i(r, \gamma, \delta) = \sum_{j=1}^{m_2} [b(\gamma, \delta) A^{-1} ]_{j,i} r_j. \tag{104} \]

By the same argument as above, we immediately deduce $F_t \alpha_i(r, \delta) - \beta(r, \gamma, \delta) = 0$ with probability one uniformly in $\Gamma$ and $\Delta$ by Assumption D. Therefore with probability one

\[ F_t(\gamma) \sum_{j=1}^{m_2} g_{j,i}(x_t, \delta) r_j - \sum_{j=1}^{m_2} [b(\gamma, \delta) A^{-1} ]_{j,i} r_j = 0, \tag{105} \]

for each $i = 1...k$, uniformly in $\Gamma$ and $\Delta$. At $\gamma = 0$, we deduce by Assumption 3

\[ F_t(0) \sum_{j=1}^{m_2} g_{j,i}(x_t, \delta) r_j = \sum_{j=1}^{m_2} [b(0, \delta) A^{-1} ]_{j,i} r_j \tag{106} \]

with probability one uniformly in $\Delta$, for every $i = 1...k$. Because the $k$-vector $r' g(x_t, \delta)$ has at least one non-degenerate element uniformly in $\Delta$ by assumption,
and the right-hand-side is a constant vector, the above equality cannot hold even at $\gamma = 0$, therefore $0 \not\in S^\ast$. Moreover, for any other $\gamma \in \Gamma$, $F_t(\gamma)$ and $F_t(\gamma) g_{t,i}(x_t, \delta)$ are $\mathcal{F}_{t-1}$-measurable with non-degenerate marginal distributions for at least one pair $i, j$ by assumption, therefore again the equality cannot hold. We deduce in this case that $S^\ast = \emptyset$.

**Case iii:** $[r'g(x_t, \delta)]_l$ is constant-valued for every $l = 1...k$, and $F_t g_{t,i,j}$ have non-degenerate marginal distributions for at least one pair $i, j$ and every $t$.

Notice that if $r'g(x_t, \delta)$ is a non-stochastic vector, then

$$\sum_{j=1}^{m_2} \left[ \beta(x_t, \delta) A^{-1} \right]_{j,i} r_j / c = r'\left[ \beta(x_t, \delta) A^{-1} \right]$$

by the definition of the matrix $A$, cf. (12). Thus, for any $r$ such that $r'g(x_t, \delta)$ is a constant $k$-vector, equality holds in (107) at the origin $\gamma = 0$. Thus, in the stochastic $g(x_t, \delta)$ case with $r'g(x_t, \delta)$ non-stochastic, the set $S^\ast = \{0\}$. □

**Proof of Corollary 9.** From the line of proof of Theorem 7, if $r'V(\gamma, \delta)r = 0$ for some $m_2$-vector $r \neq 0$, then

$$F_t \sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} f(x_t, \theta_0) \sum_{j=1}^{m_2} [g_{t,j}]_{i,j} r_j$$

$$= \sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} f(x_t, \theta_0) \sum_{j=1}^{m_2} [b A^{-1}]_{i,j} r_j.$$  

(108)

Given $\frac{\partial}{\partial \theta_i} f(x_t, \theta_0) = x_t$ and $g_t = g$, we obtain

$$F_t \sum_{i=1}^{k} x_{t,i} \sum_{j=1}^{m_2} g_{j,i} r_j$$

$$= \sum_{i=1}^{k} x_{t,i} \sum_{j=1}^{m_2} [b A^{-1}]_{j,i} r_j$$

$$F_t \sum_{i=1}^{k} x_{t,i} \alpha_i(r) = \sum_{i=1}^{k} x_{t,i} \beta_i(r, \gamma),$$

(109)

where $\alpha_i(r)$ and $\beta_i(r, \gamma)$ are defined in (96). From the line of proof of Theorem 8, we know $\beta(r, \gamma) \neq 0$ uniformly in $\Gamma$. The fact that $\beta(r, \gamma) \neq 0$ implies $\sum_{i=1}^{k} x_{t,i} \beta_i(r, \gamma) \neq 0$ with probability one due to Assumption 1: given $x_t = (y_{t-1},...,y_{t-p})'$, increasing $\sigma$-algebras $\mathcal{F}_{t-2} \subset \mathcal{F}_{t-1}$ implies $y_{t-1}$ cannot be a deterministic function of $(y_{t-2},...,y_{t-p})$.

Furthermore, because $\sum_{i=1}^{k} x_{t,i} \beta_i(r, \gamma) \neq 0$ with probability one, we deduce from equality (109) that $F_t \sum_{i=1}^{k} x_{t,i} \alpha_i(r) \neq 0$ with probability one, therefore $\sum_{i=1}^{k} x_{t,i} \alpha_i(r) \neq 0$ with probability one, hence $\alpha(r) \neq 0$.  

36
Differentiate both sides of (109) with respect to each \( x_{t,l}, l = 1...k \), multiply both sides by \( x_{t,l} \), and sum over \( l = 1...k \):

\[
F_t a_t(r) + F_t' \gamma_t \sum_{i=1}^{k} x_{t,i} a_i(r) = \beta_I(r, \gamma) \tag{110}
\]

\[
F_t \sum_{l=1}^{k} x_{t,l} a_l(r) + F_t' \sum_{i=1}^{k} \gamma_i x_{t,i} \sum_{i=1}^{k} x_{t,i} a_i(r) = \sum_{l=1}^{k} x_{t,l} \beta_I(r, \gamma).
\]

From (109) the left and right-hand side terms of (110), \( F_t \sum_{l=1}^{k} x_{t,l} a_l(r) \) and \( \sum_{l=1}^{k} x_{t,l} \beta_I(r, \gamma) \), are equal, thus cancelling renders

\[
F_t' \sum_{l=1}^{k} \gamma_l x_{t,l} \sum_{i=1}^{k} x_{t,i} a_i(r) = 0. \tag{111}
\]

Because \( \sum_{i=1}^{k} x_{t,i} a_i(r) \neq 0 \) with probability one, the zero equality holds if and only if

\[
F_t' \sum_{l=1}^{k} \gamma_l x_{t,l} = 0 \tag{112}
\]

with probability one. There are two cases to consider, \( F_t'(\gamma) \neq 0 \) or \( F_t'(\gamma) = 0 \) with probability one. If \( F_t' \neq 0 \) with probability one, then \( \sum_{l=1}^{k} \gamma_l x_{t,l} = 0 \) with probability one, which is ruled out given increasing \( \sigma \)-algebras, cf. Assumption 1, unless \( \gamma = 0 \). Thus \( \gamma = 0 \) is the only solution to the equality \( r' V(\gamma, \delta) r = 0 \), \( r \neq 0 \).

If the second case holds, \( F_t' = 0 \) with probability one, then simply differentiate both sides of (112) with respect to each \( x_{t,l} \), multiply by \( x_{t,l} \), and sum:

\[
F_t^2 \gamma_t \sum_{i=1}^{k} \gamma_i x_{t,i} + F_t' \gamma_t = 0 \tag{113}
\]

\[
F_t^2 \sum_{l=1}^{k} \gamma_l x_{t,l} \sum_{i=1}^{k} \gamma_i x_{t,i} + F_t' \sum_{l=1}^{k} \gamma_l x_{t,l} = 0.
\]

The second term \( F_t' \sum_{l=1}^{k} \gamma_l x_{t,l} \) is zero with probability one from above, hence with probability one

\[
F_t^2 \left( \sum_{l=1}^{k} \gamma_l x_{t,l} \right)^2 = 0. \tag{114}
\]

If \( F_t^2(\gamma) = 0 \) with probability, we simply repeat the differentiation, multiplication and summation performed above. Indeed, by Assumption A there exists infinitely many \( F^s(u) = (\partial/\partial u) F(u) \neq 0 \). Consider the first such \( s \): if we differentiate both sides by each \( x_{t,l} \), multiply by \( x_{t,l} \), sum over \( l = 1...k \), and note \( F_t^q \left( \sum_{l=1}^{k} \gamma_l x_{t,l} \right)^q = 0 \) for each \( F^q(u) = 0, q = 1...s - 1, \) until \( F^s(u) \neq 0 \) for some \( s \geq 1 \), we obtain

\[
F_t^s \left( \sum_{l=1}^{k} \gamma_l x_{t,l} \right)^s = 0 \tag{115}
\]

with probability one. By the arguments presented above, and because \( F_t^s \neq 0 \) with probability one, we deduce \( \gamma = 0 \) is the unique solution. \( \blacksquare \)
Proof of Theorem 10. Recall we assume \( x_t \) does not contain a constant term. For arbitrary \( \gamma \in \Gamma \) construct the function \( \tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(*)}) \) as

\[
\tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(*)}) \equiv 1 - \sum_{i=1}^{k} \hat{\gamma}_i \Psi_i(x_t) F_i'(\gamma^{(*)}) / F_i(\hat{\gamma}),
\]

where \( F'(u) \equiv (d/du)F(u) \), \( \hat{\gamma} \in \Gamma \) is arbitrary, and \( \gamma^{(*)} = (\gamma^{(1)}, ..., \gamma^{(k)}) \) is defined by (34). Define the sets \( \mathcal{S}_\gamma^{(1)} \) and \( \mathcal{S}^{(1)} \) respectively as

\[
\{ \gamma \in \Gamma : E[\xi_t \tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(*)}) F(\gamma'x_t)] = 0, \text{ and } P(\gamma'x_t \in R_0) = 1 \}
\]

and

\[
\{ \gamma \in \Gamma : E[\xi_t \tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(*)}) F(\gamma'x_t)] = 0, \text{ and } P(\gamma'x_t \in R_0) = 1 \}.
\]

We proceed in three steps: we show \( \mathcal{S}_\gamma^{(1)} \) has Lebesgue measure zero for each \( \hat{\gamma} \in \Gamma \); we then prove \( \mathcal{S}^{(1)} = \{0\} \) by proving each \( \hat{\gamma} \in \mathcal{S}_\gamma^{(1)} \), \( \hat{\gamma} \neq 0 \); finally, we prove \( \mathcal{S}^{(1)} = \{0\} \).

**Step 1:** By Assumptions 1-3 the function \( \tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(*)}) \) is uniformly bounded with probability one, and \( \hat{\gamma}, \gamma^{(*)} \in \Gamma \), a bounded subset of the reals.

Moreover, \( \tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(*)}) = 0 \) with measure zero with respect to \( \hat{\gamma} \). In order to see this, if \( \tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(*)}) = 0 \) for every \( \hat{\gamma} \) in a subset \( \Gamma_0 \) with strictly positive Lebesgue measure, then for every \( \hat{\gamma} \in \Gamma_0 \)

\[
\tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(*)}) F_i(\hat{\gamma}) \equiv F_i(\hat{\gamma}) - \sum_{i=1}^{k} \hat{\gamma}_i \Psi_i(x_t) F_i'(\gamma^{(*)}) = 0
\]

\[
F_i(\hat{\gamma}) = \sum_{i=1}^{k} \hat{\gamma}_i \Psi_i(x_t) F_i'(\gamma^{(*)}).
\]

Differentiating both sides with respect to each scalar element \( \hat{\gamma}_j \) of an arbitrary vector \( \hat{\gamma} \in \Gamma_0 \), results in for each \( j \)

\[
(\partial/\partial \hat{\gamma}_j) F_i(\hat{\gamma}) = \Psi_j(x_t) F'_i(\gamma^{(*)})
\]

\[
(\partial/\partial \hat{\gamma}_j)^s F_i(\hat{\gamma}) = 0, \ s \geq 2,
\]

with strictly positive measure, where for every \( s \geq 2 \) by construction \( (\partial/\partial \hat{\gamma}_j)^s F_i(\hat{\gamma}) \)

is identically

\[
(\partial/\partial \hat{\gamma}_j)^s F_i(\hat{\gamma}) = (\partial/\partial u)^s F(u)|_{u = \gamma' \Psi(x_t)} \times \Psi_j(x_t)^s.
\]

Thus, \( \tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(*)}) = 0 \) with strictly positive Lebesgue measure implies for each \( j \)

\[
(\partial/\partial \hat{\gamma}_j)^s F_i(\hat{\gamma}) = (\partial/\partial u)^s F(u)|_{u = \gamma' \Psi(x_t)} \times \Psi_j(x_t)^s = 0
\]

for every \( s \geq 2 \), implying \( \Psi(x_t) = 0 \) with strictly positive probability, and/or \( (\partial/\partial u)^s F(u)|_{u = \gamma' \Psi(x_t)} = 0 \) with strictly positive probability for infinite many \( s = 2, 3, ... \) Either condition contradicts Assumption A'.
Therefore, for any \( F(u) \) that satisfies Assumption \( A' \), we deduce \( \tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(s)}) = 0 \) with measure zero. The weight \( \tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(s)}) \), being uniformly bounded with probability one and non-zero with measure one, satisfies Lemma 3, hence the set \( \mathcal{S}_\gamma^{(1)} \) is countable and has Lebesgue measure zero for each point \( \hat{\gamma} \in \Gamma \).

**Step 2:** For the present step, it suffices to prove each \( \hat{\gamma} \notin \mathcal{S}_\gamma^{(1)}, \hat{\gamma} \neq 0 \): in this case,

\[
E[\epsilon_t \tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(s)})F(\hat{\gamma}'x_t)] 
eq 0 \tag{122}
\]

for arbitrary \( \hat{\gamma} \neq 0 \), hence \( \mathcal{S}^{(1)} = \{0\} \).

Recall

\[
E[\epsilon_t \Psi_i(x_t)F_i(\hat{\gamma})] \leq E[\epsilon_t \Psi_i(x_t)F_i^{(\gamma^{(i)})}] \tag{123}
\]

by construction, for every \( \hat{\gamma} \in \Gamma \).

At \( \hat{\gamma} = 0 \) notice

\[
E[\epsilon_t \tilde{h}_t^{(1)}(0, \gamma^{(s)})F_t(0)] = E[\epsilon_t F_t(0) - \epsilon_t \sum_{i=1}^{k} 0 \times \Psi_i(x_t)F_i^{(\gamma^{(i)})}] \tag{124}
\]

\[
= E[\epsilon_t F_t(0)] = cE[\epsilon_t] = 0,
\]

due to the mean-zero assumption on \( \epsilon_t \) and \( F_t(0) = c \), cf. Assumptions 1 and 3, hence 0 is an element of \( \mathcal{S}^{(1)} \). Now, separate \( \tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(s)}) \),

\[
E[\epsilon_t \tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(s)})F_t(\hat{\gamma})] = E[\epsilon_t F_t(\hat{\gamma})] - E[\epsilon_t \sum_{i=1}^{k} \hat{\gamma}_i \Psi_i(x_t)F_i^{(\gamma^{(i)})}] \tag{125}
\]

differentiate with respect to each \( \hat{\gamma}_j \), and subtract \( E[\epsilon_t \Psi_j(x_t)F_t^{(0)}] \) from both sides:

\[
\frac{\partial}{\partial \hat{\gamma}_j} E[\epsilon_t \tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(s)})F_t(\hat{\gamma})] \tag{126}
\]

\[
= E[\epsilon_t \Psi_j(x_t)F_t^{(\hat{\gamma})}] - E[\epsilon_t \Psi_j(x_t)F_t^{(\gamma^{(j)})}] \leq 0
\]

\[
\frac{\partial}{\partial \hat{\gamma}_j} E[\epsilon_t \tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(s)})F_t(\hat{\gamma})] \]

\[
= E[\epsilon_t \Psi_j(x_t) \{F_t^{(\hat{\gamma})} - F_t^{(0)}\}] - E[\epsilon_t \Psi_j(x_t) \{F_t^{(\gamma^{(j)})} - F_t^{(0)}\}] \leq 0,
\]

where the inequality holds by construction of the vectors \( \gamma^{(j)} \). Thus, the moment \( E[\epsilon_t \tilde{h}_t^{(1)}(\hat{\gamma}, \gamma^{(s)})F_t(\hat{\gamma})] \) is identically zero at the origin, \( \hat{\gamma} = 0 \), and is weakly decreasing in \( \hat{\gamma} \).

Consider two cases for any \( j \): \( E[\epsilon_t \Psi_j(x_t) \{F_t^{(\gamma^{(j)})} - F_t^{(0)}\}] = 0 \) and \( E[\epsilon_t \Psi_j(x_t) \{F_t^{(\gamma^{(j)})} - F_t^{(0)}\}] \neq 0 \). Consider the first case where the maximum centered moment satisfies \( E[\epsilon_t \Psi_j(x_t) \{F_t^{(\gamma^{(j)})} - F_t^{(0)}\}] = 0 \). The moment \( E[\epsilon_t \Psi_j(x_t) \{F_t^{(\gamma)} - F_t^{(0)}\}] \) evaluated at the origin is simply \( E[\epsilon_t \Psi_j(x_t) \{F_t^{(0)} - F_t^{(0)}\}] \). Notice that \( F'(u) - F'(0) \) satisfies Assumption \( A' \). Therefore, given \( E[\epsilon_t \Psi_j(x_t) \{F_t^{(\gamma)} - F_t^{(0)}\}] = 0 \) at \( \hat{\gamma} = 0 \), by Lemma 3 we infer there exists an open neighborhood of \( \Gamma \), say \( \mathfrak{M} \), with closure containing zero such that

39
A simple continuity argument suffices to conclude $E[\epsilon_t \Psi_j(x_t)\{F'_t(\gamma) - F'_t(0)\}] \neq 0$ for every $\gamma \in \mathfrak{N}$. Given the maximum satisfies $E[\epsilon_t \Psi_j(x_t)\{F'_t(\gamma) - F'_t(0)\}] = 0$, we deduce $E[\epsilon_t \Psi_j(x_t)\{F'_t(\gamma) - F'_t(0)\}] < 0$ for every $\gamma \in \mathfrak{N}$, $\gamma \neq 0$, and therefore strict inequality holds

$$\frac{\partial}{\partial \gamma} E[\epsilon_t \tilde{h}^{(1)}(\hat{\gamma}, \gamma^{(*)}) F_t(\hat{\gamma})]$$

for every $\gamma \in \mathfrak{N}$, $\gamma \neq 0$. We deduce the moment condition $E[\epsilon_t \tilde{h}^{(1)}(\hat{\gamma}, \gamma^{(*)}) F_t(\hat{\gamma})]$ is identically zero at the origin, is strictly decreasing arbitrarily close to the origin, and weakly decreasing everywhere else. Therefore $E[\epsilon_t \tilde{h}^{(1)}(\hat{\gamma}, \gamma^{(*)}) F_t(\hat{\gamma})] \neq 0$ for any other $\hat{\gamma} \neq 0$.

Consider the second case where the maximum centered moment is non-zero, $E[\epsilon_t \Psi_j(x_t)\{F'_t(\gamma) - F'_t(0)\}] \neq 0$. The proof is identical whether the moment is positive or negative. Assume it is positive. Because $E[\epsilon_t \Psi_j(x_t)\{F'_t(\gamma) - F'_t(0)\}] = 0$ at $\hat{\gamma} = 0$, at the origin we obtain

$$\frac{\partial}{\partial \gamma} E[\epsilon_t \tilde{h}^{(1)}(0, \gamma^{(*)}) F_t(\hat{\gamma})]$$

$$= E[\epsilon_t \Psi_j(x_t) F'_t(0)] - E[\epsilon_t \Psi_j(x_t) F'_t(\gamma^{(*)})]$$

$$= E[\epsilon_t \Psi_j(x_t) F'_t(0)] - E[\epsilon_t \Psi_j(x_t)\{F'_t(\gamma^{(*)}) - F'_t(0)\}]$$

$$= 0 - E[\epsilon_t \Psi_j(x_t)\{F'_t(\gamma^{(*)}) - F'_t(0)\}] < 0.$$

Thus, the moment condition $E[\epsilon_t \tilde{h}^{(1)}(\hat{\gamma}, \gamma^{(*)}) F_t(\hat{\gamma})]$ is identically zero at the origin, strictly decreasing at the origin and weakly decreasing everywhere else. A simple continuity argument suffices to conclude $E[\epsilon_t \tilde{h}^{(1)}(\hat{\gamma}, \gamma^{(*)}) F_t(\hat{\gamma})]$ cannot equal zero at any point other than the origin, $\hat{\gamma} = 0$. This proves $\hat{\gamma} \notin \mathfrak{S}_y^{(1)}$, $\hat{\gamma} \neq 0$, and therefore $\mathfrak{S}_y^{(1)} = \{0\}$.

**Step 3:** Under $H_1$, by Step 2 the moment condition

$$E \left[ \epsilon_t \tilde{h}^{(1)}(\gamma, \gamma^{(*)}) F_t(\gamma) \right] \neq 0$$

is guaranteed to hold for every $\gamma \neq 0$. Expanding $\tilde{h}^{(1)}(\gamma, \gamma^{(*)})$, cf. (116), we deduce under $H_1$ for every $\gamma \neq 0$

$$E \left[ \epsilon_t F_t(\gamma) \right] - \sum_{i=1}^k \gamma_i E \left[ \epsilon_t \Psi_i(x_t) F'_t(\gamma^{(i)}) \right] \neq 0. \quad (130)$$

This implies that at each point $\gamma \neq 0$ at least one moment condition, $E \left[ \epsilon_t F_t(\gamma) \right]$, $E \left[ \epsilon_t \Psi_1(x_t) F'_t(\gamma^{(1)}) \right]$, ..., $E \left[ \epsilon_t \Psi_k(x_t) F'_t(\gamma^{(k)}) \right]$ must be non-zero and therefore $E[\epsilon_t \tilde{h}^{(1)}(\gamma, \gamma^{(*)}) F_t(\gamma)] \neq 0$ for every $\gamma \neq 0$, where $\tilde{h}^{(1)}(\gamma, \gamma^{(*)})$ is defined in (33).
Moreover, expanding the vector $h_t^{(1)}(\gamma, \gamma^{(s)})$, cf. (33), and evaluating the vector moment condition $E[\epsilon_t h_t^{(1)}(\gamma, \gamma^{(s)}) F_t(\gamma)]$ at the origin, we deduce

$$E[\epsilon_t h_t^{(1)}(0, \gamma^{(s)}) F_t(0)] = 0$$

for some constant $c$, cf. (33), and evaluating the moment condition $E[\epsilon_t h_t^{(1)}(\gamma, \gamma^{(s)}) F_t(\gamma)]$ at the origin, we deduce

$$E[\epsilon_t h_t^{(1)}(0, \gamma^{(s)}) F_t(0)] = 0$$

where $E[\epsilon_t F_t(0)] = 0$ by the mean-zero assumption on $\epsilon_t$, and $F_t(0) = c$ for some constant $c$, cf. Assumptions 1 and 3. Therefore the origin $0 \in S^{(1)}$ if and only if the null innovations process $\epsilon_t$ is orthogonal to $\Psi_i(x_t) F_t(\gamma^{(i)})$, $i = 1...k$. Otherwise, $S^{(1)}$ is empty.

**Proof of Theorem 11.** The result is a direct application of Lemma 2 of Bierens (1990) and Lemma 2 of de Jong (1996). Simply replace $\epsilon_t$ in Lemma 2 of Bierens (1990) with $\epsilon_t h_t^{(1)}(\gamma, \gamma^{(s)})$. From the line of proof of Theorem 10 we know $h_t^{(1)}(\gamma, \gamma^{(s)})$ is uniformly bounded and non-zero with probability one, hence under the null we have

$$P \left( E[\epsilon_t h_t^{(1)}(\gamma, \gamma^{(s)}) | x_t] = 0 \right) = P \left( h_t^{(1)}(\gamma, \gamma^{(s)}) E[\epsilon_t | x_t] = 0 \right) = P (E[\epsilon_t | x_t] = 0) = 1,$$

and under the alternative

$$P \left( E[\epsilon_t h_t^{(1)}(\gamma, \gamma^{(s)}) | x] = 0 \right) = P (E[\epsilon_t | x_t] = 0) < 1.$$

Thus, Lemma 2 of Bierens (1990) applies to $\epsilon_t h_t^{(1)}(\gamma, \gamma^{(s)})$.

**Proof of Theorem 12.** The proof essentially mimics the line of proof of Theorem 10. Construct the $k$-vector weight $\tilde{h}_t^{(2)}(\gamma, \gamma^{(s)})$ as

$$\tilde{h}_t^{(2)}(\gamma, \gamma^{(s)}) = 1 + F_t(0)/F_t(\gamma) - \sum_{j=1}^k \gamma_j \Psi_j(x_t) F_t(\gamma^{(i)}|F_t(\gamma)) \Psi_i(x_t),$$

$i = 1...k$. Define the sets

$$\tilde{S}^{(2)}_{\gamma} = \{ \gamma \in \Gamma : E[\epsilon_t \tilde{h}_t^{(2)}(\tilde{\gamma}, \gamma^{(s)}) F_t(\gamma)] = 0, \text{ and } P(\gamma | \Psi(x_t) \in R_0) = 1 \}.$$

$$S^{(2)} = \{ \gamma \in \Gamma : E[\epsilon_t \tilde{h}_t^{(2)}(\gamma, \gamma^{(s)}) F_t(\gamma)] = 0, \text{ and } P(\gamma | \Psi(x_t) \in R_0) = 1 \}.$$

We proceed in two steps: we first prove $\tilde{S}^{(2)}_{\gamma} = \{0\}$; we then prove $S^{(2)} = \{0\}$. The proof that $\tilde{S}^{(2)}_{\gamma}$ has Lebesgue measure zero is identical to the line of proof of Theorem 9 for $S^{(1)}_{\gamma}$, hence we forego a proof here.
Step 1: Define $f(\gamma)$ as the moment condition
\[ E[\epsilon_t h_{\gamma}^{(i)}(\gamma, \gamma^{(i)}) F_i(\gamma)], \tag{136} \]
and re-write each $i^{th}$ element $f_i(\gamma)$ as
\[ f_i(\gamma) = E [\epsilon_t \Psi_i(x_t) F_i(\gamma)] - E [\epsilon_t \Psi_i(x_t) F_i(0)] - E \left[ \epsilon_t \sum_{j=1}^{k} \gamma_j \Psi_j(x_t) \Psi_j(x_t) F'_i(\gamma^{(i,j)}) \right]. \tag{137} \]
Notice that each $f_i(\gamma)$ evaluated at zero is identically zero,
\[ f_i(\gamma)|_{\gamma=0} = E [\epsilon_t \Psi_i(x_t) F_i(0)] - E [\epsilon_t \Psi_i(x_t) F_i(0)] - E \left[ \epsilon_t \sum_{j=1}^{k} 0 \times \Psi_j(x_t) \Psi_j(x_t) F'_i(\gamma^{(i,j)}) \right] = 0, \]
$i = 1..k$. Thus the origin $\gamma = 0$ is an element of each set $S_i^{(2)}$, $i = 1..k$.

Differentiating each $f_i(\gamma)$ with respect to $\gamma_j$, we deduce the inequality
\[ \frac{\partial}{\partial \gamma_j} f_i(\gamma) = E [\epsilon_t \Psi_i(x_t) \Psi_j(x_t) F'_i(\gamma)] - E \left[ \epsilon_t \Psi_i(x_t) \Psi_j(x_t) F'_i(\gamma^{(i,j)}) \right] \leq 0 \tag{139} \]
by the construction of $\gamma^{(i,j)}$. At the origin $\gamma = 0$ we deduce by Assumption 3 for some scalar constant $\tilde{c}$ the $j^{th}$-gradient component satisfies
\[ \frac{\partial}{\partial \gamma_j} f_i(\gamma)|_{\gamma=0} = E [\epsilon_t \Psi_i(x_t) \Psi_j(x_t) F'_i(0)] - E \left[ \epsilon_t \Psi_i(x_t) \Psi_j(x_t) F'_i(\gamma^{(i,j)}) \right] = \tilde{c} E [\epsilon_t \Psi_i(x_t) \Psi_j(x_t)] - E \left[ \epsilon_t \Psi_i(x_t) \Psi_j(x_t) F'_i(\gamma^{(i,j)}) \right] = E \left[ \epsilon_t \Psi_i(x_t) \Psi_j(x_t) \{ \tilde{c} - F'_i(\gamma^{(i,j)}) \} \right] = E \left[ \epsilon_t \Psi_i(x_t) \Psi_j(x_t) \tilde{F}'_i(0, \gamma^{(i,j)}) \right] \leq 0 \tag{140} \]
where we define $\tilde{F}'_i(\gamma, \gamma^{(i,j)}) = F'_i(\gamma) - F'_i(\gamma^{(i,j)})$, and the inequality follows by construction.

Consider two cases: for any $i, j$, \[ E[\epsilon_t \Psi_i(x_t) \Psi_j(x_t) \tilde{F}'_i(0, \gamma^{(i,j)})] = 0, \tag{141} \]
or \[ E[\epsilon_t \Psi_i(x_t) \Psi_j(x_t) \tilde{F}'_i(0, \gamma^{(i,j)})] \neq 0. \tag{142} \]
Consider the first case, and observe that $\tilde{F}'(\gamma, \gamma^{(i,j)})$ satisfies Assumption $A'$. Hence, by Lemma 3 because $E[\epsilon_t \Psi_i(x_t) \Psi_j(x_t) F'_i(0, \gamma^{(i,j)})] = 0$ there exists a neighborhood of $\Gamma$, say say $\mathcal{N}$, with closure containing zero such that
\[ E \left[ \epsilon_t \Psi_i(x_t) \Psi_j(x_t) \tilde{F}'_i(\gamma, \gamma^{(i,j)}) \right] < 0 \tag{143} \]
for each $\gamma \in \mathfrak{N}$, where the strict inequality follows from the boundedness
\[ E[\epsilon_t \Psi_i(x_t) \Psi_j(x_t) \hat{F}_i'(\gamma^{(i,j)})] \leq 0 \tag{144} \]
by construction. Thus, for $\gamma$ arbitrarily close to zero the gradient component $\partial/\partial \gamma_i f_i(\gamma)$ identically satisfies
\[ \frac{\partial}{\partial \gamma_i} f_i(\gamma) = E[\epsilon_t \Psi_i(x_t) \Psi_j(x_t) F_i'(\gamma)] - E \left[ \epsilon_t \Psi_i(x_t) \Psi_j(x_t) F_i'(\gamma^{(i,j)}) \right] < 0. \tag{145} \]
We deduce $f_i(\gamma) = 0$ at the origin $\gamma = 0$, the slope $\partial/\partial \gamma_j f_i(\gamma) < 0$ for every $j$ arbitrarily close to the origin, and $\partial/\partial \gamma_j f_i(\gamma) \leq 0$ everywhere else. Therefore the $i$th moment condition $f_i(\gamma)$ cannot equal zero at any point other than the origin.

For the second case, $E[\epsilon_t \Psi_i(x_t) \Psi_j(x_t) \tilde{F}_i'(\gamma^{(i,j)})] < 0$. Because for all nuisance vectors $\partial/\partial \gamma_j f_i(\gamma) \leq 0$, and at the origin $f_i(0) = 0$ and
\[ \frac{\partial}{\partial \gamma_j} f_i(\gamma)|_{\gamma=0} = E \left[ \epsilon_t \Psi_i(x_t) \Psi_j(x_t) \tilde{F}_i'(\gamma^{(i,j)}) \right] < 0 \tag{146} \]
we immediately deduce again that the $i$th moment condition $f_i(\gamma) \neq 0$ for all non-zero $\gamma$.

**Step 2:** The remainder of the proof mimics the line of proof of Step 3 of Theorem 10 with $E[\epsilon_t F_i(\gamma)]$ replaced by $E[\epsilon_t \Psi_i(x_t) F_i(\gamma)]$, $i = 1...k$, and $E[\epsilon_t \Psi_i(x_t) F_i'(\gamma^{(i,j)})]$ replaced by $E[\epsilon_t \Psi_i(x_t) \Psi_j(x_t) F'(\gamma^{(i,j)}/F_x)]$, $i = 1...k$, $j = i...k$. \[ \square \]

**Corollary 13.** The result is a direct application of Theorem 8. Recall we assume $f(x_t, \theta) = \theta' x_t$ and $\Psi(x_t) = x_t$. Each $h_{t,i,j}^{(2)}(\gamma, \gamma^{(s)})$ is therefore defined as
\[ h_{t,i,j}^{(2)}(\gamma, \gamma^{(s)}) = x_t x_t \tilde{F}_i'(\gamma^{(i,j)/F_x})/F'(\gamma'/F_x), \tag{147} \]
and $h_{t}^{(2)}(\gamma, \gamma^{(s)})$ stacks $x_t F_i(0)/F_0(\gamma)$, $x_t$ and each $h_{t,i,j}^{(1)}(\gamma, \gamma^{(s)})$, $1 \leq i \leq j \leq k$. The vector weight $h_{t}^{(2)}(\gamma, \gamma^{(s)})$ may, therefore, be written as
\[ h_{t}^{(2)}(\gamma, \gamma^{(s)}) = g(x_t, \delta)x_t \tag{148} \]
where the \(k/2(k + 3) \times k\) matrix \(g(x_t, \delta)\) is constructed as

\[
\begin{bmatrix}
F_0^t(0) & 0 & \ldots & 0 \\
0 & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & 0 \\
0 & \ldots & 0 & F_0^t(0) \\
1 & 0 & \ldots & 0 \\
0 & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & 0 \\
0 & \ldots & 0 & 1 \\
\end{bmatrix}
\]

The second upper \(k \times k\) block of \(g(x_t, \delta)\) is simply the identity matrix, thus \(g(x_t, \delta)\) is guaranteed to be non-zero. Moreover, by Assumption 1 and Assumption 3, \(F^t_i(\gamma^{(i,j)})\) and \(F_t(\gamma)\) are \(\mathfrak{F}_{i-1}\)-measurable, and by Assumption 1 \(x_t\) is \(\mathfrak{F}_{i-1}\)-measurable with non-degenerate marginal distributions. Therefore, at least one element \(g_{t,j}(x_t, \delta)\) has a non-degenerate marginal distribution.

However, notice that there exist some \(m_2\)-vectors \(r\) such that \(r'g(x_t, \delta)\) has only degenerate marginal distributions. For example, if \(r_i = 1\) for each \(i = k + 1 \ldots 2k\), and \(r_i = 0\) for all other \(i\), then \(r'g(x_t, \delta) = I_k\). Therefore Theorem 7.iii applies and \(S^*(2) = \{0\}\).

### References


