Consistent Testing of Functional Form in Time Series Models

James Davidson and Andreea G. Halunga
Department of Economics
University of Exeter Business School
Rennes Drive
Exeter EX4 4PU
U.K.

Abstract

We develop a consistent procedure for testing the adequacy of parametric time series models. The approach is to extend Herman Bierens' idea of examining the covariances between regression residuals and an exponential weight function, to check the full range of orthogonalities predicted for the score contributions in quasi-maximum likelihood estimation. Tests of this type, which involve nuisance parameters, are defined as either 'sup'ed or integrated conditional moment tests, and are often implemented using bootstrap methods. However, our emphasis in this study is on practical implementation. We study a two-statistic approach that aims to exploit the available power while keeping computing requirements to a minimum.

1 Introduction

In this chapter, we study some tests for consistent model specification in time series models. In particular, we are interested in methods that have power to detect arbitrary deviations from the null hypothesis, which may itself be a nonlinear time series model. The central idea is to construct a test of an infinite number of moment restrictions by computing the covariance of a suitably defined target series with a function of test variables admitting an infinite series expansion. The exponential function is a natural choice. Pioneering work has been done on these methods, using the regression residuals as the target series, by Bierens (1984, 1987, 1988, 1990), de Jong (1996), and Bierens and Ploberger (1997), inter alia.

A number of extensions to Bieren's work have been considered in recent literature. For example, Hill (2005) extends Bierens' and de Jong's approach to construct a consistent test that has maximal power against smooth transition autoregressive (STAR) alternatives. Kasparis (2010) extends Bierens test in the context of non-stationary regressors. Whang (2000, 2001) and Delgado, Dominguez and Lavergne (2006) propose consistent tests in an i.i.d. context using an indicator function instead of the exponential weighting function of Bierens. Escanciano (2007) provides a unified theory for both continuous and discontinuous weighting functions using 'residual marked' empirical processes, to detect misspecifications in time series regression models. In semiparametric dynamic models, Chen and Fan (1999) extend the Bierens (1990) approach to testing conditional moment restrictions using the weighted integrated squared metric.

All these tests focus on regression model residuals, and accordingly test specification of the conditional mean. Escanciano (2007) and, in the non-parametric framework, Hsiao and Li (2001) propose consistent tests for conditional heteroskedasticity under the null of homoskedasticity. However, they do not suggest a consistent test for the correct specification of the GARCH(1,1) null model. The latter authors give some indication that such tests could be constructed using the framework of de Jong (1996). De Jong himself considers an increasing number of lags as the sample size increases when constructing the weighting function. This has the shortcoming that it requires numerical integration with dimension of the order of the sample size, which makes this test effectively infeasible in large samples.

Another approach to consistent tests of functional form is to compare the fitted parametric regression function with a nonparametric model. Within the framework of independently and identically distributed observations, such tests have been proposed by Zheng (1996), Eubank and Spiegelman (1990), Härdle and Mammen (1993), Hong and White (1996), Fan and Li (1996a), inter alia. For time series, such developments include Fan and Li (1996b) and Hsiao and Li (2001). Hong (1993) and Zheng (1994, 1996) propose consistent tests for heteroskedasticity under the null hypothesis of i.i.d. for both the regressand and regressors. An extension to conditional heteroskedasticity within the time series framework has been proposed by Hsiao and Li (2001). Although these tests are consistent against all alternatives to the null hypothesis, they require smoothing of the data and have nontrivial power only under local alternatives that approach the null at a rate slower than the square root of the sample size.

In Davidson and Halunga (2012) we consider a natural extension of the moment testing principle in the context of maximum likelihood (or quasi-maximum likelihood) estimation. This is to test the full range of hypothesized orthogonalities implied by the model, specifically, choosing as target process the vectors of score contributions. Our aim in the present paper is to extend these approaches to consider tests of dynamic specification. We develop tests appropriate for routine use in econometric modelling practice and pay particular attention to issues of computational economy. Some proposed test procedures entail a great deal of calculation, involving numerical integration or optimization over nuisance parameters combined with the use of the bootstrap to generate p-values. Our view is that tests that take more than a few seconds to evaluate on a standard computer system are likely to be neglected by practitioners, and compromises may be necessary to meet these requirements. We focus our attention on refining the 'two-statistic' trick suggested in Bierens (1990), exploiting an optimized test statistic while allowing the use of chi-squared critical values.

The paper is organized as follows. Section 2 sets out the theoretical background and describes the procedures to be investigated. Section 3 states the assumptions underlying the asymptotic analysis of the tests and proves the main results, showing the null distribution and the consistency property. Section 4 provides additional details of our implementation of the tests and reports the results of a range of simulation experiments. Section 5 concludes the paper, and an appendix contains proofs.

2 Tests of dynamic specification based on score contributions

Correct dynamic specification of a model of a time series process y_t is characteristically defined by the (un)predictability of certain functions of data and parameters. Given a filtration $\mathcal{F}_t = \sigma(y_s, x_s, s \leq t\}$ where x_t is a vector of 'conditioning variables', consider a vector $d_t(\theta)$ of $\mathcal{F}_t$-measurable random functions where $\theta \in \Theta \subseteq \mathbb{R}^p$. Letting $\mathcal{I}_t = \sigma(y_s, s < t; x_s, s \leq t\} \subseteq \mathcal{F}_t$, the null hypothesis of correct specification can be stated formally as the existence of $\theta_0 \in \Theta$ such that

$$P\left(E\left[d_t\left(\theta_0\right)|\mathcal{I}_t\right] = 0\right) = 1 \text{ for } t = 1, .., T \tag{2.1}$$

with the alternative hypothesis

$$P\left(E\left[d_t\left(\theta\right)|\mathcal{I}_t\right] = 0\right) < 1, \text{ for all } \theta \in \Theta \text{ and at least one } t. \tag{2.2}$$

In the literature on consistent testing, d_t is customarily a regression residual. Consider instead the idea of basing a test on the p-vector of score contributions. In other words, define a quasi-log likelihood function as

$$L_T = \sum\nolimits_{t=1}^{T} l_t \tag{2.3}$$

and choose $d_t = \partial l_t/\partial\theta$ $(p \times 1)$. The quasi-maximum likelihood estimator is $\hat{\theta} = \arg\min_{\theta\in\Theta} L_T$, such that $\sum_{t=1}^{T} d_t(\hat{\theta}) = 0$. Under the hypothesis of correct specification, $E(l_t|\mathcal{I}_t)$ is maximized at θ_0 with probability 1 for each t, equivalent (subject to second order conditions) to (2.1). We emphasize 'quasi' here because l_t is not required to be the true log-density of the data. Since $\mathcal{F}_t = \sigma(y_t, \mathcal{I}_t)$ and $\mathcal{F}_{t-1} \subseteq \mathcal{I}_t$, the null hypothesis includes the condition that the score contributions have the martingale difference property but may also imply a further orthogonality restriction relating to 'exogenous' variables x_t, contemporaneously dated but treated as causally prior, to y_t. There's no implication here that l_t depends on all the elements of x_t under H_0, and our test is focused equally on incorrect functional form and omitted effects.

Consider the standard case of the regression model with possible conditional heteroscedasticity,

$$y_t = m_t(\theta) + h_t(\theta)^{1/2}\varepsilon_t$$

where m_t and h_t are $\mathcal{I}_t$-measurable random functions with $h_t = \sigma^2$ in the standard homoscedastic case, and by hypothesis $E(\varepsilon_t|\mathcal{I}_t) = 0$ and $E(\varepsilon_t^2|\mathcal{I}_t) = 1$. As is well-known, setting

$$l_t = \log h_t + \frac{(y_t - m_t)^2}{h_t} \tag{2.4}$$

yields consistent and asymptotically normal QML estimators under a range of mild regularity conditions, and in this case

$$d_t = 2\frac{(y_t - m_t)}{h_t}\frac{\partial m_t}{\partial\theta} + \left(\frac{(y_t - m_t)^2}{h_t} - 1\right)\frac{1}{h_t}\frac{\partial h_t}{\partial\theta}. \tag{2.5}$$

The orthogonality conditions to be satisfied under the null hypothesis relate closely to the usual conditions on the residuals $y_t - m_t$ and their squares. However, (2.1) represents the full set of orthogonalities predicted by the null hypothesis in this model, and it is natural to test them jointly. A p-degree of freedom test can be based on this multiple restriction, or elements of d_t can be tested individually. Note that the principle of testing the scores extends to cases where unique vectors of residuals do not naturally arise, such as discrete data models and Markov-switching models.

The natural way to test H_0 is by examining the sample covariances of the target process with an $\mathcal{I}_t$-measurable test function. Let $w_t(\xi)$ denote such a function, typically constructed as a bounded scalar, where ξ is a vector of nuisance parameters falling in a compact set $\Xi \subset \mathbb{R}^K$ to be chosen by the investigator. Conditional M-tests can be constructed on the generic indicator

$$s_T(\hat{\theta}, \xi) = \frac{1}{T}\sum\nolimits_{t=1}^{T} d_t(\hat{\theta}) w_t(\xi) \quad (p \times 1). \tag{2.6}$$

The basic CM test statistic takes the form

$$S_T(\xi) = T s_T(\hat{\theta}, \xi)' \hat{V}_T(\xi)^{-1} s_T(\hat{\theta}, \xi) \tag{2.7}$$

where $\hat{V}_T(\xi)$ is an estimator of the asymptotic covariance matrix

$$V(\xi) = R(\xi) - Q(\xi) M^{-1} P(\xi)' - P(\xi) M^{-1} Q(\xi)' + Q(\xi) M^{-1}\Sigma M^{-1} Q(\xi)'. \tag{2.8}$$

Here, M, Σ, $Q(\xi)$, $P(\xi)$ and $R(\xi)$ are the limits as $T \to \infty$ of the matrices of expectations

$$M_T = \frac{1}{T}\sum_{t=1}^{T} E\left[-\partial d_t(\theta)/\partial\theta'\right]_{\theta=\theta_0} \tag{2.9}$$

$$\Sigma_T \;=\; \frac{1}{T}\sum_{t=1}^{T} E\left[d_t(\theta)d_t(\theta)'\right]_{\theta=\theta_0}, \tag{2.10}$$

and for each ξ,

$$Q_T(\xi) = \frac{1}{T}\sum_{t=1}^{T} E\left[-w_t(\xi)\frac{\partial d_t(\theta)}{\partial\theta'}\right]_{\theta=\theta_0} \tag{2.11}$$

$$P_T(\xi) = \frac{1}{T}\sum_{t=1}^{T} E\left[w_t(\xi)d_t(\theta)d_t(\theta)'\right]_{\theta=\theta_0} \tag{2.12}$$

$$R_T(\xi) = \frac{1}{T}\sum_{t=1}^{T} E\left[w_t(\xi)^2 d_t(\theta)d_t(\theta)'\right]_{\theta=\theta_0}. \tag{2.13}$$

Letting $\hat{M}_T$, $\hat{\Sigma}_T$, $\hat{P}_T(\xi)$, $\hat{Q}_T(\xi)$ and $\hat{R}_T(\xi)$ denote the variants of formulae (2.9)-(2.13) with expectations replaced by realized values evaluated at the consistent estimator $\hat{\theta}$, $\hat{V}_T(\xi)$ is constructed by letting these matrices replace their limiting counterparts in (2.8). Under standard regularity conditions, $\sqrt{T}s_T(\hat{\theta},\xi) \xrightarrow{d} N(0, V(\xi))$ pointwise in Ξ and hence $S_T(\xi) \to_d \chi^2(p)$ when the null hypothesis is true.

Following Bierens (1990) and de Jong (1996) *inter alia*, the chosen weight function has the form

$$w_t(\xi) = \exp\left\{\xi'\psi(\tilde{z}_t)\right\} \tag{2.14}$$

where z_t $(K \times 1)$ denotes the vector of $\mathcal{I}_t$-measurable test variables, $\tilde{z}_t$ represents the vector standardized to have mean 0 and variance 1, and $\psi : \mathbb{R}^K \mapsto (-\pi/2, \pi/2)^K$ denotes the arctangent function, ensuring that the arguments are bounded and possess all their moments. The exponential is a natural choice of function to ensure that the test covariance involves an infinite set of co-moments of the test variables and residuals, although any function having an infinite series expansion could in principle be substituted (Stinchcombe and White, 1998).

The weight function must perform the exacting duty of capturing the dependencies of $\mathcal{I}_t$-measurable variables with the target series, and our choice of target series is motivated to provide the best chance for this to happen. Considering the case of (2.5), compare our approach with the usual choice of $u_t = y_t - m_t$ as target series. One set of functions to be tested for correlation with u_t have the form $(\partial m_t/\partial\theta)w_t/h_t$. The null hypothesis requires the orthogonality of u_t with each element of $\partial m_t/\partial\theta$, and the sample correlations with w_t should indicate the failure of any of these conditions. When, as is normally the case, $\partial m_t/\partial\theta$ has a constant element corresponding to an intercept, the original indicator based on u_t itself is included in the test set. Similar considerations apply to the test functions associated with $u_t^2/h_t - 1$.

It can be shown (see Lemma 3.1) that the set of points of Ξ for which the resulting test is inconsistent has Lebesgue measure zero. Therefore, the strategy of picking ξ at random from a uniform distribution on Ξ would suffice for consistency. However, optimal power considerations, in addition to the desire for a single reproducible procedure, call for some form of averaging or optimizing of the statistic with respect to ξ.

One can think of ξ as a vector of parameters defining a pseudo-alternative hypothesis and, accordingly, not identified under the null hypothesis. A number of papers, notably Bierens (1990), Bierens and Ploberger (1997), Andrews and Ploberger (1994) and Hansen (1996) have considered the problem of eliminating dependence on such parameters by integrating them out with respect to an auxiliary distribution, defining a class of integrated conditional moment (ICM) tests. Letting $S_T(\xi)$ denote the conditional M-statistic, define

$$\hat{S}_T^A = \int_{\Xi} S_T(\xi)d\xi \tag{2.15a}$$

$$\hat{S}_T^B = 2\log\int_\Xi \exp\left\{\tfrac{1}{2}S_T(\xi)\right\}d\xi \tag{2.15b}$$

$$\hat{S}_T^S = \sup_{\xi\in\Xi} S_T(\xi). \tag{2.15c}$$

(Where we wish to discuss these alternatives collectively without distinguishing between them, we henceforth write simply $\hat{S}_T$.) These statistics are special cases of the class of tests described by Andrews and Ploberger. $\hat{S}_T^A$ is motivated by those authors as conferring best power for local alternatives close to the null. (We think here in terms of local alternatives represented by $\theta_0 = (\theta_{10}', T^{-1/2}\delta')'$ where $\hat{\theta} = (\hat{\theta}_1, 0')'$ represents the estimator subject to the restrictions of the null hypothesis.) Simple averaging is a reasonable strategy when the statistic $S_T(\xi)$ does not vary greatly over Ξ. In the case of $\hat{S}_T^B$, the statistic places more weight in the average on the larger values of $S_T(\xi)$, and is said by Andrews and Ploberger to confer best power in larger cases of the local alternative δ. The sup statistic, proposed originally by Davies (1977, 1987) can be viewed as the limit of $2r^{-1}\log\int_\Xi \exp\left\{\frac{r}{2}S_T(\xi)\right\}d\xi$ as $r \to \infty$, and is said to be optimal for local alternatives furthest from the null. All the tests in (2.15) can therefore be termed integrated conditional moment (ICM) tests. Bierens (1982) and De Jong (1996) suggest analytic evaluation of the ICM integrals, although in large samples the computational overhead can be non-trivial. For the present work we have used Monte Carlo integration, which allows a flexible trade-off between cost and numerical accuracy.

The difficulty with the ICM statistics in applications is that their null distributions are dependent on the generation process of the data, and hence are not amenable to tabulation. Their distribution might be approximated by a direct Monte Carlo algorithm like that proposed by Hansen (1996), but this still makes for a very computationally intensive procedure requiring integral evaluation at each Monte Carlo replication. A more practical approach is the one suggested by Bierens (1990), in which the ICM statistic is computed just once, together with a case with fixed $\xi = \xi_0$, whose asymptotic null distribution is known. Consider the test statistic

$$\tilde{S}_T = \begin{cases} S_T(\xi_0), & \hat{S}_T - S_T(\xi_0) \le \gamma T^\rho \\ \hat{S}_T, & \hat{S}_T - S_T(\xi_0) > \gamma T^\rho \end{cases} \tag{2.16}$$

for suitably chosen constants γ and $0 < \rho < 1$. Under the null hypothesis the difference $\hat{S}_T - S_T(\xi_0) = O_p(1)$, like the statistics themselves, and $\tilde{S}_T = S_T(\xi_0)$ with probability converging to 1 as T increases. Under the alternative hypothesis, either both statistics are $O_p(T)$, or $\hat{S}_T - S_T(\xi_0) = O_p(T)$ and $\tilde{S}_T = \hat{S}_T$ with probability converging to 1 as T increases. With γ and ρ chosen large enough, this scheme allows us to assert the known asymptotic distribution when the null hypothesis is true, and at the same time, if γ and ρ are not too big, gain additional power under the alternative. We call this the 'two-statistic' test. While its behaviour in large samples is known for arbitrary choices of γ and ρ, "large" here may need to be interpreted as very large indeed. The behaviour of the procedure in samples of moderate size, and the critical choice of γ in particular, is examined by simulation in Section 4.

In the dynamic context, we face a major difficulty with the construction of truly consistent tests, since in a consistent test the set of test variables needs to include all observed lags of the relevant series. Truncating the length of lag arbitrarily carries with it the possibility of failing to detect a specific departure from the null. However, the absence of any natural basis for truncating the lags at a finite point carries the equally disagreeable implication that the number of nuisance parameters is large and increasing with sample size. This is the scheme investigated by de Jong (1996), but the computational burden associated with his procedure proves to be very severe. Within our feasibility constraints, consistency in this literal sense appears difficult to achieve.

Some other way must be found to deal with lag distributions, than assigning each lagged variable its own weight.

Bierens (1988) offers an elegant argument to show (assuming a univariate process for convenience of notation only) that

$$E(y_t|\sigma(y_{t-1}, y_{t-2}, y_{t-3}, \ldots)) = E(y_t|\sigma(\textstyle\sum_{j=1}^{\infty} \tau^{j-1} y_{t-j}))$$

for all τ in a set of Lebesgue measure 1 from $(-1, 1)$. To ensure that the mapping from $y_{t-1}, y_{t-2}, y_{t-3}, \ldots$ to $\sum_{j=1}^{\infty} \tau^{j-1} y_{t-j}$ is Borel-measurable, Bierens' proof requires only that the data series be rational-valued, a trivial requirement in practice. However, since the moving average would need to be stored with an arbitrarily large number of decimal digits for the information it contained to be practically retrievable, this result is more in the nature of a possibility theorem than a practical tool.

Let the dynamic test function have the basic form

$$w_t(\xi) = \exp\left\{\sum\nolimits_{j=0}^{c-1} \lambda_j' \psi(\tilde{r}_{t-j})\right\}. \tag{2.17}$$

We consider two approaches to including lagged variables in a feasible manner. The first is to truncate the lag distributions at a finite point, so that $z_t = (r_t', r_{t-1}', \ldots, r_{t-c+1}')'$ where r_t $(m \times 1)$ is $\mathcal{I}_t$-measurable, $\xi = \left(\lambda_0', ..., \lambda_{c-1}'\right)'$ $(cm \times 1)$ and the λ_j are fixed m-vectors, so $K = cm$. In this case the first c observations of the estimation sample will in this case be lost. The second approach is to let $c = t$ in (2.17) so that $z_t = (r_t', r_{t-1}', \ldots, r_1')'$, but to invoke an assumption of smoothness of the lag distributions in the manner of Almon (1965), letting

$$\lambda_j = \sum\nolimits_{i=1}^{P} (j+1)^{-i-1} \xi_i \tag{2.18}$$

so that $\xi = \left(\xi_1', ..., \xi_p'\right)'$ and $K = Pm$. The weights are required to decline at least at the rate $O(j^{-2})$ to ensure summability. This option increases the dimension of Ξ only by a factor P, which can be chosen moderately large so that a variety of lag structures can be economically approximated.

While the indicated specification of the function w_t ensures that it is technically bounded for all choices of data set, the question of scale variations is still crucial to the performance of these tests. To take one extreme case, note that we were to choose $w_t = 1$, then $s_T = 0$ identically and $M = Q$ and $R = P = \Sigma$ in (2.8). With too little scale variation the limit distribution is undefined. On the other hand, under the exponential transformation even a technically bounded argument could exhibit variations wide enough that the convergence of higher moments would be slow and the Gaussian approximation correspondingly poor. Ensuring a suitable range of variation is an important practical issue. Our approach is to normalize $\xi'\psi(\tilde{z}_t)$ to have a fixed range independent of the variation of the underlying data. We set this range in our simulation experiments at 3. In other words, we set the range of variation of so that $0.22 \leq w_t \leq 4.48$ in all tests. In effect, this means that the boundaries of Ξ are being set dynamically to match the range of variation of $\psi(\tilde{z}_t)$. Note however that the asymptotic analysis in Section 3 does not assume this refinement, and accordingly covers a larger set of cases.

In addition to the test of joint restrictions, there are various other ways of extracting information from the indicator to yield consistent tests. In particular, we construct one degree of freedom tests based on the elements of the test vector. The basic test statistics are constructed as

$$S_i(\xi) = \frac{T^{-1}\left(\sum_{t=1}^{T} d_{ti}(\hat{\theta}) w_t(\xi)\right)^2}{\{\hat{V}_T(\xi)\}_{ii}} \tag{2.19}$$

where $d_{ti}(\hat{\theta}) = \partial l_t/\partial\theta_i|_{\theta=\hat{\theta}}$, for $i = 1, ..., p$, and $\{\hat{V}_T(\xi)\}_{ii}$ is the ith diagonal element of $\hat{V}_T(\xi)$. Individual tests are then defined for $i = 1, \ldots, p$ by computing any of the functionals (2.15) for the cases $S_i(\xi)$.

3 Asymptotic Analysis

In this section we formally derive the asymptotic properties of the tests under specified assumptions. Since the definition of the dynamic weight functions involves variable truncations and hence an implicit array framework, it is convenient to consider the case where the lags are extended to infinity. Accordingly we write

$$w_t^\infty(\xi) = \exp\left\{\sum\nolimits_{j=0}^{\infty} \lambda_j'\psi(\tilde{z}_{t-j})\right\} \tag{3.1}$$

where it is understood that the λ_j' may either be zero for $j \geq c < \infty$ so that $w_t^\infty = w_t$ trivially, or are otherwise subject to polynomial smoothness constraints and hence dependent on at most P parameters, also forming an absolutely summable sequence at all points of Ξ. Expressions depending on w_t may be similarly decorated to indicate the substitution of w_t^∞.

Assumptions

1. The observed data $(y_t', x_t')'$, $t = 1, .., T$ form a sequence of strictly stationary and ergodic random variables.
2. The parameter space Θ is a compact subset of $\mathbb{R}^p$.
3. $d_t(\theta) : \mathbb{R}^{G+K} \times \Theta \longmapsto \mathbb{R}^p$ is a Borel measurable function for each $\theta \in \Theta$ and continuously differentiable on Θ.
4. For all t and some $s > 0$, the following are bounded uniformly in t,[1]

   (i) $E\left[\sup_{\theta\in\Theta} \|d_t(\theta)\|^{2(1+s)}\right]$,

   (ii) $E\left[\sup_{\theta\in\Theta,\xi\in\Xi} \|d_t(\theta)w_t^\infty(\xi)\|^{2(1+s)}\right]$,

   (iii) $E\left[\sup_{\theta\in\Theta} \left\|\partial d_t(\theta)/\partial\theta'\right\|^{1+s}\right]$,

   (iv) $E\left[\sup_{\theta\in\Theta,\xi\in\Xi} \left\|\partial d_t(\theta)/\partial\theta' w_t^\infty(\xi)\right\|^{1+s}\right]$.
5. $M = \lim_{T\to\infty} M_T$ defined in (2.9) is finite and non-singular;
6. Under the null hypothesis (2.1),

   (i) $d_t(\theta_0)$ is a vector martingale difference,

   (ii) $\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, M^{-1}\Sigma M^{-1}\right)$, where $\Sigma = \lim_{T\to\infty} \Sigma_T$ defined in (2.10).

These assumptions are deliberately set at a high level in respect of time series properties, specifying required behaviour of objects such as $d_t(\theta_0)$ and $\hat{\theta}$, rather than specific conditions on the data series sufficient for them to hold. Such conditions are now well-known. Where appropriate we assume that the functions of the data and parameters arising in the sequel preserve stationarity and ergodicity, and that restrictions on the parameter space are such as to ensure this requirement.

The following lemmas establish the basis for the consistent test.

[1] Throughout the paper, $\|\cdot\|$ denotes the Euclidean norm of a vector or matrix.

Lemma 3.1 *If $P\left(E\left[d_t(\theta)|\mathcal{F}_{t-1}\right]=0\right)<1$, then for any $\theta\in\Theta$ the set*

$$B_0=\left\{\xi\in\Xi:\ E\left[d_t\left(\theta\right)w_t^{\infty}\left(\xi\right)\right]=0\right\}$$

has Lebesgue measure zero.

Lemma 3.2 *Under Assumptions 1-6 and H_0 (2.1),*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}d_t(\hat{\theta})w_t(\xi)\xrightarrow{d}N\left(0,V(\xi)\right)$$

pointwise in Ξ.

Assumption 7 The set $B^*=\{\xi\in\Xi:\operatorname{rank}\left(V(\xi)\right)<p\}$ has Lebesgue measure zero.

Subject to Assumption 7, (2.7) provides a consistent specification test. $V\left(\xi\right)$ should have rank p under the same circumstances that Σ has rank p for all ξ except on a set of Lebesgue measure zero. The asymptotic distribution of (2.7) for given ξ is established as follows.

Theorem 3.1 *For every $\xi\in\Xi-B_0\cup B^*$, where B_0 is the set defined in Lemma 3.1 for the case $\theta=\theta_0$, and B^* is the set defined in Assumption 7, under H_0 in (2.1) $S_T\left(\xi\right)\rightarrow_d\chi^2(p)$, whereas under H_1 in (2.2), $S_T\left(\xi\right)/T\rightarrow q\left(\xi\right)$ a.s., where $q\left(\xi\right)>0$.*

Next consider the statistics $\hat{S}_T$ defined by (2.15). Let C_Ξ denote the metric space of real continuous functions endowed with the uniform metric $\sup_{\xi\in\Xi}\left\|z_1\left(\xi\right)-z_2\left(\xi\right)\right\|$. Since all three statistics are continuous functionals with domain C_Ξ, we can treat them in comparable fashion, by application of the following result.

Theorem 3.2 *Under H_0 and Assumptions 1-7, $\sqrt{T}s_T(\hat{\theta},\xi)$ defined in (2.6) converges weakly to a mean-zero Gaussian element $z\left(\xi\right)$ of C_Ξ with covariance function*

$$E\left[z\left(\xi_1\right)z\left(\xi_2\right)'\right]=V\left(\xi_{1,}\xi_2\right)$$

where

$$V\left(\xi_1,\xi_2\right)=R\left(\xi_1,\xi_2\right)-Q\left(\xi_1\right)M^{-1}P\left(\xi_2\right)'-P\left(\xi_1\right)M^{-1}Q\left(\xi_2\right)'+Q\left(\xi_1\right)M^{-1}\Sigma M^{-1}Q\left(\xi_2\right)'$$

and $R\left(\xi_1,\xi_2\right)=\lim T^{-1}\sum_{t=1}^{T}E\left[d_t(\theta)d_t(\theta)'w_t^{\infty}\left(\xi_1\right)w_t^{\infty}\left(\xi_2\right)\right]_{\theta=\theta_0}$, $R\left(\xi,\xi\right)=R\left(\xi\right).$

Let $G(\cdot)$ denote any of the three continuous functionals defined for $S_T(\xi)$ in (2.15), so that $G(S_T)$ denotes one of $\hat{S}_T^A$, $\hat{S}_T^B$ and $\hat{S}_T^S$. Since the set $B_0\cup B^*$ has Lebesgue measure zero, it follows by the continuous mapping theorem that under H_0

$$G(S_T)\xrightarrow{d}G(z\left(\xi\right)'V\left(\xi\right)^{-1}z\left(\xi\right)).$$

Finally, following Bierens (1990) note the following, letting $G(\cdot)$ be defined as above.

Theorem 3.3 *Under Assumptions 1-7, let $\tilde{S}_T$ be defined by (2.16). Under H_0, $\tilde{S}_T\rightarrow_d\chi^2(\ p)$, whereas under H_1, $\tilde{S}_T/T\rightarrow\max(q(\xi_0),G(q(\xi))>0$ a.s. as $T\rightarrow\infty$.*

4 Monte Carlo Evidence

This section reports Monte Carlo experiments on the performance of the proposed tests in a set of linear and nonlinear univariate time series models. In all these experiments, 10,000 replications were performed to estimate rejection frequencies under the null hypothesis. 5000 replications were performed on most cases of the alternative hypothesis, where the estimation of tail probabilities is less critical.

In all cases the models are estimated by Gaussian (quasi-) maximum likelihood. Statistics of the form $\tilde{S}^A$ and $\tilde{S}^B$ in (2.15) are computed by Monte Carlo integration, evaluating the statistic at repeated random drawings from the uniform distribution on the set Ξ and cumulating these outputs until the convergence criterion is met, here chosen as the absolute effect of an additional draw on the average falling below 0.002. The set Ξ is defined as the K-dimensional hypercube with upper bounds 1 and lower bounds -1, although in view of the re-scaling of the weights noted in Section 2, this choice is essentially arbitrary.

To compute the 'sup' statistic is somewhat trickier since it is important that the optimization algorithm should perform independently of starting values, and should be able to handle arbitrary functions with no assumption of differentiability or continuity. The method adopted is to make uniform random drawings from a K-dimensional region, initially Ξ. At each iteration the $S_T(\xi)$ statistics are evaluated at the drawn points, ranked, and the smallest two-thirds of the cases discarded. The search region is then contracted to the smallest hypercube containing the remainder, before adding new draws from this region to the set. The convergence criterion is that both the diameter of the search set and the range of statistic values in the set falls below 0.002. A maximum of 5000 statistic evaluations is imposed in all the procedures. Under reasonable smoothness assumptions, note that highly accurate evaluation of these integrated statistics is not essential.

The questions we seek to study here are of a practical nature. A key issue is the choice of bound in the implementation of the 'two-statistic' test. The baseline statistic $S_T(\xi_0)$ is computed with $\xi_0 = (1, \ldots, 1)'$. Optimally, γ and ρ need to be as small as is compatible with keeping the probability of the bound being exceeded to a low level when the null hypothesis is true. Two data-independent characteristics of the test likely to affect the distribution of $\hat{S}_T - S_T(\xi_0)$ are the degrees of freedom of the statistic (p) and the dimension of Ξ (K). We experimented with a range of settings, and one thing that became evident is that in the case of $\tilde{S}_T^S$ the distribution depends markedly on whether $K = 1$ or $K > 1$. In the former case there is only a modest scope for altering the statistic through the nonlinear mapping from scalar z_t to w_t, but much greater scope exists when ξ is a vector. The value of K matters much less for the $\tilde{S}_T^A$ and $\tilde{S}_T^B$ statistics, however. A range of formulations have been studied, and the general form finally adopted is

$$Bound = \gamma_0 T^{\rho_1} D^{\rho_2} (1 + I(\text{sup test}, \ K > 1) K^{\rho_3})$$

where T denotes sample size, D denotes the degrees of freedom of the test, $I(\cdot)$ is the indicator function of its arguments and γ_0, ρ_1, ρ_2, and ρ_3 are constants to be selected. In all the experiments reported, the parameters used are $\gamma_0 = 2.5$, $\rho_1 = 0.2$, $\rho_2 = 0.4$ and $\rho_3 = 0.1$. We do not report the experiments that lead to these choices, but present evidence (see Table 1) showing that they do a reasonable job of optimizing the procedure.

For each of our models, we first computed for comparison a residual-based consistent test in the manner of Bierens (1990) but with one of our proposed weighting functions and integrated variants. This is denoted $\tilde{B}$ in the tables. The joint score-based test in (2.7), having p degrees of freedom, is denoted $\tilde{S}$, and the individual tests on score elements are then denoted $\hat{S}_\theta$ for the various cases of θ, as defined in (2.19). Each of these symbols may refer in the tables either to the 'sup' version or the exponential ICM version of the test, respectively, as the context will indicate.

(We do not report any experiments with the case $\tilde{S}^A$ in this paper.)

Finally, three cases of the weight function w_t are compared. These are the polynomial lag with $P = 1$, the polynomial lag with $P = 3$, and the free inclusion of the three leading lags. Since our examples are univariate, this means that $K = P$ in the first two cases, and $K = 3$ in the third one. In the tables, these alternatives are labelled 'Poly-1', 'Poly-3' and '3 lags', respectively. Given the nature of the alternatives actually simulated, it is to be expected that the truncated lag case does no worse than the others. The object is to establish how far there is a cost to choosing the procedure that is nominally consistent against general alternatives over the more flexible one.

Consider Table 1, where we present evidence on the 'two-statistics' procedure. The model being simulated here is the AR(1),

$$y_t = 1 + 0.5y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1). \tag{4.1}$$

The table shows the performance of the six variants of the tests, as described, for three sample sizes. The two tests computed here are the usual Bierens test of the residuals ($D = 1$) and the joint score contributions-based test. Here $D = 3$, the parameters being the intercept ϕ_0, autoregressive coefficient ϕ_1 and residual variance σ^2.

Each cell of the table has three entries. The first (Roman font) is the estimated test size expressed as the percentage of rejections of the (true) null. The second entry (sloping font) is the percentage of the replications in which *Excess* > *Bound*, where '*Excess*' here denotes $\hat{S}_T - S_{0T}$, or $\hat{B}_T - B_{0T}$ as the case may be. For correctly sized tests based on chi-squared critical values, this indicator needs to be either zero or very close to it. The maximum attained in these experiments is 0.26% and the average is under 0.1%. The third entry (typewriter font) is the ratio to *Bound* of the maximum value of *Excess* achieved over the replications. This should be as close to 1 as possible, or power will be sacrificed unnecessarily. The smallest value observed in these experiments is 0.69, but there does not appear to be a discernable pattern that would suggest modifying our choice of bound parameters. The evidence of this table shows that our choices of *Bound* work well, at least for this simple model.

Table 2 shows the performance of the same tests in the context of estimation of the AR(1) model when the true data generation process has the ESTAR form

$$y_t = 1 + 0.5y_{t-1} \exp\left(-0.4y_{t-2}^2\right) + \varepsilon_t \quad \varepsilon_t \sim N(0, 1). \tag{4.2}$$

This table shows additionally the individual tests based on each score element. In this table, the percentage of cases where the bound was exceeded appear in square brackets following the rejection percentages. Don't overlook the fact that these values depend on the choice of ξ_0, so the comparisons across the different cases are more informative than the percentages themselves.

Next, Table 3 shows the result of testing in the context of over-fitted models. Autoregressions of order 1, 2, 3 and 4 have been fitted with 500 observations, where the true model is ESTAR as in (4.2). In addition, the performance of some conventional diagnostic tests are reported: the Ljung-Box (1978) and McLeod-Li (1983) tests based on the autocorrelations of residuals and the squared residuals, each with 12 lags; the Breusch-Godfrey LM test for autocorrelation (Breusch 1978, Godfrey 1978) and the Engle (1982) LM test for ARCH each with 4 lags specified, and finally Ramsey's (1969) RESET test based on the squared fitted values.

The idea behind this experiment is to throw light on the common modelling strategy of choosing lag lengths in the light of residual autocorrelation tests. How does this strategy perform when the true model is nonlinear? Two conclusions can be drawn. First, we note that the autocorrelation tests have no power to detect mis-specification in an over-fitted model. The ESTAR model involves second order lags, whose omission in the AR(1) is detected by the autocorrelation statistics, but the linear approximation represented by the AR(2) serves to effectively suppress residual autocorrelation. The McLeod-Li and ARCH tests perform poorly, showing that squared residuals are a poor proxy for the missing components. By contrast, the consistent tests retain a good measure of power in the over-fitted cases. The striking feature of this table is the performance of the individual score-based test corresponding to the second lag, which shows the greatest rejection rate in all the over-fitted cases. This finding illustrates the claimed virtue of the score-based approach, of pinpointing the source of the mis-specification. Here, in the AR(p) cases for $p \geq 2$, there is clear evidence that the second lag is problematic, while there is no issue with the other parameters. While the tests cannot point us directly at the correct specification, they at least offer a clue.

The remaining tables illustrate the performance of the tests in a range of different models, both cases of the null hypothesis and of alternatives. To check null rejection frequencies Table 4 reports the results of simulating and then estimating the following models, where $\varepsilon_t \sim N(0,1)$ in each case.

ARMA: $y_t = 1 + 0.7y_{t-1} + \varepsilon_t + 0.3\varepsilon_{t-1}$

AR2: $y_t = 1 + 0.5y_{t-1} + 0.3y_{t-2} + \varepsilon_t$

GARCH1: $y_t = \sqrt{h_t}\varepsilon_t, \quad h_t = 0.05 + 0.1y_{t-1}^2 + 0.4h_{t-1}$

AR-GARCH: $y_t = 1 + 0.5y_{t-1} + u_t, \quad u_t = \sqrt{h_t}\varepsilon_t, \quad h_t = 0.05 + 0.1u_{t-1}^2 + 0.4h_{t-1}.$

The tests reported here are ICM-B tests with three lags. In the case of the AR-GARCH model a proportion of the replications reported convergence failure of the optimization algorithm. These samples were discarded and the number of replications extended. The reported results relate to the successful estimations only.

Table 5 reports the percentage of rejections in a range of nonlinear alternatives. The fitted model is in every case the AR(1), and in this case the test is the ‘sup’ version of the ‘3 lags’ test. The models simulated, in addition to those previously defined, are as follows where $\varepsilon_t \sim N(0,1)$ in every case.

SETAR: $y_t = 1 + 0.5y_{t-1} + (-1 - 0.9y_{t-1})\, I(y_{t-1} > 1) + \varepsilon_t$

SGN: $y_t = \operatorname{sgn}(y_{t-1}) + \varepsilon_t$

BILIN: $y_t = 1 + 0.5y_{t-1} + 0.7y_{t-1}\varepsilon_{t-1} + \varepsilon_t$

NLMA: $y_t = 1 + 0.5y_{t-1} + 0.3\varepsilon_{t-1}\varepsilon_{t-2} + \varepsilon_t$

MARKOV-SW: $y_t = \left\{\begin{matrix} 4 + 0.3y_{t-1} + \varepsilon_t \text{ if } & S_t = 1 \\ 1 - 0.5y_{t-1} + \varepsilon_t \text{ if } & S_t = 2 \end{matrix}\right\}$ with transition probabilities

$$P(S_t = j | S_{t-1} = k) = \left\{ \begin{matrix} 0.9, & j = k \\ 0.1\ , & j \neq k \end{matrix} \right\}\ j, k = 1, 2.$$

ARCH: $y_t = 1 + 0.5y_{t-1} + u_t, \quad u_t = \sqrt{h_t}\varepsilon_t, \qquad h_t = 0.05 + 0.8u_{t-1}^2$

GARCH2: $y_t = 1 + 0.5y_{t-1} + u_t, \ u_t = \sqrt{h_t}\varepsilon_t, \ h_t = 0.05 + 0.4u_{t-1}^2 + 0.4h_{t-1}.$

The most interesting feature of this table is the quite wide range of test powers in evidence. While detection of some alternatives is very effective, with others it is not. In particular, the low rate of detection of conditional heteroskedasticity (the ARCH and GARCH2 cases) is quite surprising, and this case deserves further investigation, notwithstanding that good tests for these effects do of course exist. However, in the context of GARCH estimation the tests are powerful in detecting functional form mis-specification. In Table 6, the estimated model is GARCH(1,1) (whose performance under the null is shown in Table 4) and the data generation process is one of the following GARCH variants:

$$\begin{aligned}&\text{GJR: } y_t = 1 + u_t, \qquad u_t = \sqrt{h_t}\varepsilon_t\\ &\qquad h_t = 0.005 + 0.17\,|u_{t-1}|\,(1 + 1.5I\,(u_{t-1} < 0)) + 0.6h_{t-1}\end{aligned}$$

$$\begin{aligned}&\text{EGARCH: } y_t = 1 + u_t, \qquad u_t = \sqrt{h_t}\varepsilon_t\\ &\qquad \ln(h_t) = 0.005 + 0.17\left|u_{t-1}/\sqrt{h_{t-1}}\right|\left(1 + 1.5I\left(u_{t-1}/\sqrt{h_{t-1}} < 0\right)\right) + 0.6\ln(h_{t-1}).\end{aligned}$$

The tests work particularly well in discriminating between GARCH and EGARCH. In this table, the GARCH intercept is denoted by γ and the coefficients by α and β as in the usual Bollerslev (1987) notation

Finally Table 7 provides evidence on the effectiveness of the infinite lag options in constructing the weight functions. While most of the models tested here are naturally suited to the truncated lag option, the ARMA(1,1) is AR(∞) and so constructing the weight functions with an infinite lag of the measured series is nominally appropriate. The results show that a polynomial order of at least 6 is required for a powerful test in this case. The same number of free lags generally does a little better, so there is evidently quite a fine trade-off between truncation and constraining the form of the lag distribution.

5 Concluding Remarks

Monte Carlo evidence on a selection of models is inevitably anecdotal, and in this paper we have not attempted a comprehensive coverage of the several combinations of test options available. Nonetheless, some provisional conclusions emerge quite clearly. The tests are generally well sized, even in the smaller samples according to Tables 1 and 4. The 1 degree of freedom tests tend if anything to be undersized, which suggests that the deviation from the asymptotic distribution of S_{0T} makes a larger contribution to the error in rejection probability than the bounds device. The tests do however have excellent power against a range of alternatives, although the frequency with which the optimized statistic dominates the baseline case – such that it is the value reported – is not always large. In other words, the baseline test can work well in its own right, in the cases considered.

There is not much to choose in terms of rejection rates between the 'sup' test and the ICM-B test. (The ICM-A test has not been studied). The regular Bierens test on residuals generally works well, and enjoys the benefit of being a 1-degree of freedom test and hence suffers a smaller penalty in the two-statistics setup. While we expect the weight function based on a small number of unrestricted lags to perform best against the alternatives considered, it is still a surprise to note in Tables 2 and 7 that the 'Poly-3' weight function based on a third-order polynomial generally does worse than 'Poly-1', although in Table 7 the 'Poly-6' cases does better than either. The most reasonable explanation for this finding is the additional penalty imposed on the bound for the case $K > 1$. However, reference to Table 1 indicates that this penalty cannot be dispensed without compromising good size characteristics. The 'free lag' options turn out to be the best in spite of this handicap, reminding us that consistency against all alternatives must involve a compromise. We have evidence of a fairly delicate trade-off involved in the specification of these tests.

References

[1] Bierens, H.J., 1982. Consistent model specification tests. *Journal of Econometrics* 20, 105-134.

[2] Bierens. H.J., 1984. Model specification testing of time series regressions, *Journal of Econometrics* 26, 323-353.

[3] Bierens. H.J., 1987. ARMAX model specification testing, with an application to unemployment in the Netherlands. *Journal of Econometrics* 35, 161-191.

[4] Bierens. H.J.. 1988, Arma memory index modeling of economic time series, *Econometric Theory* 4. 35-59.

[5] Bierens, H.J., 1990. A consistent conditional moment test of functional form, *Econometrica* 58, 1443-1458.

[6] Bierens, H.J., 1991. Least squares estimation of linear and nonlinear armax models under data heterogeneity. *Annales d'Economie et de Statistique* 20/21, 143-169.

[7] Bierens, H.J., and Ploberger, W., 1997. Asymptotic theory of integrated conditional moment tests, *Econometrica* 65, 1129-1151.

[8] Bollerslev, T., 1986. Generalized autoregressive heteroskedasticity, *Journal of Econometrics*, 31, 307-27.

[9] Bollerslev, T. 1988: On the correlation structure of the generalized autoregressive conditional heteroscedastic process. *Journal of Time Series Analysis* 9, 121–31.

[10] Breusch, T. S. 1978: Testing for autocorrelation in dynamic linear models. *Australian Economic Papers* 17, 334–55.

[11] Chen, X., and Fan, Y., 1999. Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series, *Journal of Econometrics* 91, 373-401.

[12] Davidson, J., and A. Halunga (2012) Consistent model specification tests. Working paper, University of Exeter.

[13] de Jong, R.M., 1996. The Bierens test under data dependence. *Journal of Econometrics* 72, 1-32.

[14] Delgado, M.A., Dominguez, M.A., and Lavergne, P., 2006. Consistent tests of conditional moment restrictions. *Annales d'Economie et de Statistique* 81, 33-67.

[15] Engle, R.F., 1982 Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica*, 50, 987-1007.

[16] Escanciano, J.C., 2007. Model checks using residual marked empirical processes. *Statistica Sinica* 17, 115-138.

[17] Eubank, R.L., and Spielgelman, C.H., 1990. Testing the goodness of fit of a linear model via nonparametric regression techniques, *Journal of the American Statistical Association* 85, 387-392.

[18] Fan, Y., and Li, Q., 1996a. Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica* 64, 865-890.

[19] Fan, Y., and Li, Q., 1996b. Consistent model specification tests: kernel-based tests versus Bierens' ICM tests. Unpublished manuscript. Department of Economics, University of Windsor.

[20] Godfrey, L. G. 1978: Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica* 46, 1293–302.

[21] Hall, P., and C. C. Heyde, 1980. *Martingale Limit Theory and its Applications*, New York, Academic Press.

[22] Härdle, W., and Mammen, E., 1993. Comparing nonparametric versus parametric regression fits, *Annals of Statistics* 21, 1926-1947.

[23] Hill, J. B., 2008. Consistent and non-degenerate model specification tests against smooth transition and neural networks alternatives. *Annales D'Economie et de Statistique* 90, 145-179.

[24] Hong, Y., and White, H., 1996. Consistent specification testing via nonparametric series regressions. *Econometrica* 63, 1133-1160.

[25] Hsiao, C., and Q. Li, 2001 A consistent test for conditional heteroskedasticity in time series regression models, *Econometric Theory* 17, 188-221

[26] Kasparis, I., 2010. The Bierens test for certain nonstationary models, *Journal of Econometrics*, 158, p. 221-230.

[27] Ling. S. and M. McAleer, 2003. Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* 19, 280-310.

[28] Ljung, G. M., and G. E. P. Box 1978. On a measure of lack of fit in time-series models, *Biometrika*, 65, 297-303.

[29] McLeod, A. I. and Li,W. K. 1983 Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis* 4, 269-273.

[30] Newey, W.K., 1985. Maximum likelihood specification testing and conditional moment test. *Econometrica* 53, 1047-1070.

[31] Newey, W.K., 1991. Uniform convergence in probability and stochastic equicontinuity, *Econometrica* 59, 1161-1167

[32] Ramsey, J. B. 1969: Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society, Series B* 31, 350–71.

[33] Stinchcombe, M.B., and White, H., 1998. Consistent specification testing with nuisance parameters present only under the alternative, *Econometric Theory* 14, 295-325.

[34] Tauchen, G., 1985. Diagnostic testing and evaluation of maximum likelihood models, *Journal of Econometrics* 30, 415-443.

[35] White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test of heteroskedasticity, *Econometrica* 48, 817-38.

[36] Whang, Y-J., 2000. Consistent bootstrap tests of parametric regression functions, *Journal of Econometrics* 98, 27-46.

[37] Whang, Y-J , 2001. Consistent specification testing for conditional moment restrictions. *Economics Letters* 71, 299-306.

[38] White, H., 1982. Maximum likelihood estimation of misspecified models, *Econometrica* 50, 1-26.

[39] Zheng, J.X., 1996. A consistent test of functional form via nonparametric estimation techniques, *Journal of Econometrics* 75, 263-289

A Appendix

Proof of Lemma 3.1. This follows from Lemma 1 of Bierens (1990) under the alternative hypothesis in (2.2).

Lemma A.1 *Consider $\xi = (\xi_1, \ldots, \xi_K)' \in [-b, b]^K$ for $b \leq 1$, and random vector $q = (q_1, \ldots, q_K)'$ with support $[-a, a]^K$. If $\|\xi_1 - \xi_2\| \leq \tau$ then*

$$\left|\exp(\xi_1' q) - \exp(\xi_2' q)\right| \leq \tau b^{K-1} \exp(Ka)$$

holds with probability 1.

Proof. For arbitrary sets of numbers $a_{11}, \ldots, a_{1K}$ and $a_{21}, \ldots, a_{2K}$, applying the convention $\prod_{p=m}^{n} a_{up} = 1$ if $m > n$ for $u = 1, 2$,

$$\begin{aligned} |a_{11} \cdots a_{1K} - a_{21} \cdots a_{2K}| &= \left| \sum_{j=1}^{K} (a_{1j} - a_{2j}) \prod_{p=1}^{j-1} a_{1p} \prod_{p=j+1}^{K} a_{2p} \right| \\ &\leq \sum_{j=1}^{K} |a_{1j} - a_{2j}| \prod_{p=1}^{j-1} |a_{1p}| \prod_{p=j+1}^{K} |a_{2p}| \end{aligned} \tag{A-1}$$

Applying the multinomial expansions of the terms in the power series representation of the exponentials yields

$$\begin{aligned} &\left| \exp\left(\sum_{k=1}^{K} \xi_{1k}' q_k \right) - \exp\left(\sum_{k=1}^{K} \xi_{2k}' q_k \right) \right| \\ &\qquad \left| \sum_{i=0}^{\infty} \frac{1}{i!} \left[\left(\sum_{k=1}^{K} \xi_{1k}' q_k \right)^i - \left(\sum_{k=1}^{K} \xi_{2k}' q_k \right)^i \right] \right| \\ &\qquad = \left| \sum_{i=0}^{\infty} \frac{1}{i!} \sum_{k_1, \ldots, k_K = 0}^{i} \binom{i}{k_1, \ldots, k_K} \left(\xi_{11}^{k_1} \cdots \xi_{1K}^{k_K} - \xi_{21}^{k_1} \cdots \xi_{2K}^{k_K} \right) q_1^{k_1} \cdots q_K^{k_K} \right| \\ &\qquad \leq \tau b^{K-1} \sum_{i=0}^{\infty} \frac{1}{i!} \sum_{k_1, \ldots, k_K = 0}^{i} \binom{i}{k_1, \ldots, k_K} \left| q_1^{k_1} \cdots q_K^{k_K} \right| \\ &\qquad = \tau b^{K-1} \exp\left(\sum_{k=1}^{K} |q_k| \right) \\ &\qquad \leq \tau b^{K-1} \exp(Ka) \end{aligned} \tag{A-2}$$

where the first inequality follows from (A-1), putting $a_{uj} = \xi_{uj}^{k_j}$ and noting that the majorant is bounded by τb^{K-1}, by assumption. ∎

Lemma A.2 *If*

$$g_u(j) = \sum_{k=1}^{K} \xi_{uk} q_k(j)$$

for $u = 1, 2$, where for each k, $\{q_k(j)\}$ is a sequence of random variables with support $[-a, a]$, $|\xi_{uk}| \leq b \leq 1$ and $\|\xi_1 - \xi_2\| \leq \tau$, then

$$\left| \exp\left(\textstyle\sum_{j=0}^{\infty} (1+j)^{-2} g_1(j) \right) - \exp\left(\textstyle\sum_{j=0}^{\infty} (1+j)^{-2} g_2(j) \right) \right| \leq C\tau$$

holds with probability 1, where

$$C = b^{K-1}\zeta(2(K-1))\exp\left(Ka(1+2b\zeta(2))\right.$$

and $\zeta(\cdot)$ *denotes the Riemann zeta function.*

Proof.

$$\begin{aligned}
&\left|\exp\left(\sum_{j=0}^{\infty}(1+j)^{-2}g_1(j)\right)-\exp\left(\sum_{j=0}^{\infty}(1+j)^{-2}g_2(j)\right)\right| \\
&\qquad\leq \sum_{j=0}^{\infty}\Bigg[\big|\big(\exp((1+j)^{-2}g_1(j)\big)-\big(\exp((1+j)^{-2}g_2(j)\big)\big| \\
&\qquad\qquad\times\exp\left(\sum_{l=0}^{j-1}(1+l)^{-2}g_1(l)\right)\exp\left(\sum_{l=j+1}^{\infty}(1+l)^{-2}g_2(l)\right)\Bigg] \\
&\qquad\leq \exp\left(\zeta(2)Kab\right)\sum_{j=0}^{\infty}\big|\big(\exp((1+j)^{-2}g_1(j)\big)-\big(\exp((1+j)^{-2}g_2(j)\big)\big| \\
&\qquad\leq \exp\left(\zeta(2)Kab\right)\tau b^{K-1}\sum_{j=0}^{\infty}(1+j)^{-2(K-1)}\exp\left(\sum_{k=1}^{K}|q_k(j)|\right) \\
&\qquad\leq \exp\left(\zeta(2)Kab\right)\tau b^{K-1}\zeta(2(K-1))\exp\left(Ka\right) \qquad\text{(A-3)}
\end{aligned}$$

where the first inequality in (A-3) follows by (A-1) applied to the bounded infinite products (that is, infinite sums under the exponential). The third inequality applies Lemma A.1 term by term, noting how the coefficients are in effect bounded absolutely by $b(j+1)^{-2} \leq 1$, and hence the sequence of bounds is summable. ▮

Lemma A.3 *Under Assumptions 1-4*

$$\sup_{\theta\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^{T}d_t(\theta)d_t(\theta)'-\lim_{T\to\infty}E\left[\frac{1}{T}\sum_{t=1}^{T}d_t(\theta)d_t(\theta)'\right]\right\| = o_p(1) \qquad\text{(A-4)}$$

$$\sup_{\theta\in\Theta,\xi\in\Xi}\left\|\frac{1}{T}\sum_{t=1}^{T}w_t(\xi)\,d_t(\theta)-\lim_{T\to\infty}E\left[\frac{1}{T}\sum_{t=1}^{T}w_t^{\infty}(\xi)\,d_t(\theta)\right]\right\| = o_p(1) \qquad\text{(A-5)}$$

$$\sup_{\theta\in\Theta,\xi\in\Xi}\left\|\frac{1}{T}\sum_{t=1}^{T}w_t(\xi)\,d_t(\theta)d_t(\theta)'-\lim_{T\to\infty}E\left[\frac{1}{T}\sum_{t=1}^{T}w_t^{\infty}(\xi)\,d_t(\theta)d_t(\theta)'\right]\right\| = o_p(1) \qquad\text{(A-6)}$$

$$\sup_{\theta\in\Theta,\xi\in\Xi}\left\|\frac{1}{T}\sum_{t=1}^{T}(w_t(\xi))^2\,d_t(\theta)d_t(\theta)'-\lim_{T\to\infty}E\left[\frac{1}{T}\sum_{t=1}^{T}(w_t^{\infty}(\xi))^2\,d_t(\theta)d_t(\theta)'\right]\right\| = o_p(1) \qquad\text{(A-7)}$$

$$\sup_{\theta\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial d_t(\theta)}{\partial\theta'}-\lim_{T\to\infty}E\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial d_t(\theta)}{\partial\theta'}\right]\right\| = o_p(1) \qquad\text{(A-8)}$$

$$\sup_{\theta\in\Theta,\xi\in\Xi}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(w_t(\xi)\,\frac{\partial d_t(\theta)}{\partial\theta'}\right)-\lim_{T\to\infty}E\left[\frac{1}{T}\sum_{t=1}^{T}\left(w_t^{\infty}(\xi)\,\frac{\partial d_t(\theta)}{\partial\theta'}\right)\right]\right\| = o_p(1) \qquad\text{(A-9)}$$

Proof. Firstly, (A-4) and (A-8) follow applying a uniform law of large numbers (ULLN) for strictly stationary and ergodic processes (e.g., see Ling and McAleer (2003), Theorem 3.1). For a generic function $q_t(\theta)$, to show that

$$\sup_{\theta\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^{T}q_t(\theta)-\lim_{T\to\infty}E\left(\frac{1}{T}\sum_{t=1}^{T}q_t(\theta)\right)\right\|=o_p(1)$$

it is sufficient to establish that $E\sup_{\theta\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^{T}q_t(\theta)\right\|^{1+s}<\infty$ uniformly in t for some $s>0$. The condition follows by the Cauchy-Schwarz inequality and Assumption 4(i), and Assumption 4(iii), respectively.

The other components of the lemma are established as follows. Again for generic q_t, let

$$r_T(\theta,\xi)=\frac{1}{T}\sum_{t=1}^{T}w_t(\xi)\,q_t(\theta).$$

and define $r_T^{\infty}(\theta,\xi)$ similarly with $w_t^{\infty}(\xi)$ replacing $w_t(\xi)$. Note that

$$\begin{aligned}\sup_{\theta\in\Theta,\xi\in\Xi}\|r_T(\theta,\xi)-E[r_T^{\infty}(\theta,\xi)]\| &\le \sup_{\xi\in\Xi,\theta\in\Theta}\|r_T^{\infty}(\theta,\xi)-E[r_T^{\infty}(\theta,\xi)]\|\\ &\quad+\sup_{\xi\in\Xi,\theta\in\Theta}\|r_T^{\infty}(\theta,\xi)-r_T(\theta,\xi)\| \qquad \text{(A-10)}\end{aligned}$$

and it suffices to show that the two right-hand side terms of (A-10) are $o_p(1)$.

For the first term, the ULLN for strictly stationary and ergodic processes may be applied as above, given Assumption 4. For the second term, note that we can write

$$w_t(\xi)=\exp\left(\sum_{j=0}^{t-1}(1+j)^{-2}g(j)\right)$$

where, with λ_j defined by (2.18), $g(j)=(1+j)^2\lambda_j'\psi(\tilde r_{t-j})$ is a sequence of almost surely bounded random variables, and $w_t^{\infty}(\xi)$ has the corresponding representation. If $\Xi=[-b,b]^K$ then $|g(j)|<KbP\pi/2$ with probability 1. Since there exists $C<\infty$ such that $\left|\sum_{j=t}^{\infty}(1+j)^{-2}g(j)\right|<Ct^{-1}$ almost surely, it follows that

$$\begin{aligned}|w_t^{\infty}(\xi)-w_t(\xi)| &= \exp\left(\sum_{j=0}^{t-1}(1+j)^{-2}g(j)\right)\left|\exp\left(\sum_{j=t}^{\infty}(1+j)^{-2}g(j)\right)-1\right|\\ &\le \exp\left(\sum_{j=0}^{t-1}(1+j)^{-2}g(j)\right)\left|\exp(Ct^{-1})-1\right|\\ &= O(t^{-1}). \qquad \text{(A-11)}\end{aligned}$$

Therefore,

$$\begin{aligned}\sup_{\theta,\xi}\|r_T^{\infty}(\theta,\xi)-r_T(\theta,\xi)\| &\le \sup_{\theta}T^{-1}\sum_{t=1}^{T}\|r_t(\theta)\|\sup_{\xi}|w_t^{\infty}(\xi)-w_t(\xi)|\\ &= O_p(T^{-1}\log T)\\ &= o_p(1) \quad \blacksquare \qquad \text{(A-12)}\end{aligned}$$

Proof of Lemma 3.2. The proof consists in establishing the following steps for fixed $\xi \in \mathbb{R}^K$:

(i) $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sup_{\theta\in\Theta} \|d_t(\theta) w_t^{\infty}(\xi) - d_t(\theta) w_t(\xi)\| = o_p(1)$;

(ii) $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} d_t(\hat{\theta}) w_t(\xi) = \sqrt{T} z_T(\theta_0,\xi) + o_p(1)$ where

$$z_T(\theta_0,\xi) = \frac{1}{T}\sum_{t=1}^{T} d_t(\theta_0) w_t^{\infty}(\xi) - Q_T^{\infty}(\xi) M_T^{-1} \frac{1}{T}\sum_{t=1}^{T} d_t(\theta_0) \tag{A-13}$$

and $Q_T(\xi)$ and M_T are defined in (2.11) and (2.9) respectively, with Q_T^{∞} denoting the substitution of w_t^{∞} for w_t in Q_T.

(iii)

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} d_t(\theta_0) w_t^{\infty}(\xi) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} d_t(\theta_0) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} R(\xi) & P(\xi) \\ P(\xi)' & \Sigma \end{pmatrix} \right),$$

where $R(\xi)$, $P(\xi)$ and Σ are the respective limits as $T \to \infty$ of $R_T(\xi)$, $P_T(\xi)$ and Σ_T defined in (2.13), (2.12) and (2.10).

Step (i), follows by (A-11) and an arguments similar to (A-12). Step (ii) follows using part (i), consistency of $\hat{\theta}$, and a mean value expansion of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} d_t(\hat{\theta}) w_t^{\infty}(\xi)$ about the true parameter θ_0. Finally step (iii) is established by applying a CLT for martingale differences (e.g, Corollary 3.2 of Hall and Heyde (1980)). ▮

Lemma A.4 *Under H_0 and Assumptions 1-6,*

$$\hat{V}(\xi)^{-1/2}\sqrt{T} s_T(\hat{\theta},\xi) - V(\xi)^{-1/2}\sqrt{T} z_T(\theta_0,\xi) = o_p(1) \tag{A-14}$$

uniformly over $\xi \in \Xi$, where $z_T(\theta_0,\xi)$ is defined in (A-13).

Proof. We have that

$$\begin{aligned} &\sup_{\xi\in\Xi} \left\| \hat{V}(\xi)^{-1/2}\sqrt{T} s_T(\hat{\theta},\xi) - V(\xi)^{-1/2}\sqrt{T} z_T(\theta_0,\xi) \right\| \\ &\quad \le \sup_{\xi\in\Xi} \left\| \hat{V}(\xi)^{-1/2} - V(\xi)^{-1/2} \right\| \sup_{\xi\in\Xi} \left\| \sqrt{T} s_T(\hat{\theta},\xi) \right\| \\ &\quad + \sup_{\xi\in\Xi} \left\| \sqrt{T} s_T(\hat{\theta},\xi) - \sqrt{T} z_T(\theta_0,\xi) \right\| \sup_{\xi\in\Xi} \left\| V(\xi)^{-1/2} \right\| \end{aligned} \tag{A-15}$$

By Lemmas 3.2 and A.3, and Slutsky's Theorem

$$\sup_{\xi\in\Xi} \left\| \hat{V}(\xi)^{-1/2} - V(\xi)^{-1/2} \right\| = o_p(1).$$

Moreover, by Lemma 3.2

$$\begin{aligned} \sqrt{T} s_T(\hat{\theta},\xi) &= \sqrt{T} z_T(\theta_0,\xi) + o_p(1) \\ &= O_p(1) \end{aligned}$$

uniformly over ξ. Therefore,

$$\sup_{\xi\in\Xi} \left\| \hat{V}(\xi)^{-1/2} - V(\xi)^{-1/2} \right\| \sup_{\xi\in\Xi} \left\| \sqrt{T} s_T(\hat{\theta},\xi) \right\| = o_p(1).$$

Now

$$\sup_{\xi\in\Xi}\left\|\sqrt{T}s_T(\hat{\theta},\xi)-\sqrt{T}z_T\left(\theta_0,\xi\right)\right\|=o_p\left(1\right)$$

by Lemma 3.2 and since $\sup_{\xi\in\Xi}\left\|V(\xi)^{-1/2}\right\|=O_p\left(1\right)$, the second term in the expression (A-15) is $o_p\left(1\right)$. ∎

Lemma A.5 *Under Assumptions 1-7 and H_1, there exists for each $\xi\in\Xi$ a function $\pi_\xi:\mathbb{R}^p\rightarrow\mathbb{R}^p$ such that*

$$\hat{V}(\xi)^{-1/2}s_T(\hat{\theta},\xi)-V(\xi)^{-1/2}\pi_\xi=o_p\left(1\right)$$

where $V(\xi)^{-1/2}\pi_\xi\neq 0$ for all $\xi\in\Xi$ except possibly in a set of Lebesgue measure zero.

Proof. We can write for each $\xi\in\Xi$

$$\begin{aligned}\left\|\hat{V}(\xi)^{-1/2}s_T(\hat{\theta},\xi)-V(\xi)^{-1/2}\pi_\xi\right\|&\leq\left\|\hat{V}(\xi)^{-1/2}-V(\xi)^{-1/2}\right\|\left\|\pi_\xi\right\|\\&+\left\|s_T(\hat{\theta},\xi)-\pi_\xi\right\|\left\|\hat{V}(\xi)^{-1/2}\right\|.\end{aligned}\tag{A-16}$$

For the second right-hand side term, $\hat{V}(\xi)^{-1/2}=O_p\left(1\right)$ and (A-5) of Lemma A.3 establishes that

$$\operatorname*{plim}_{T\to\infty}\sup_{\theta\in\Theta}\left\|s_T\left(\theta,\xi\right)-\lim_{T\to\infty}E\left[s_T\left(\theta,\xi\right)\right]\right\|=0.$$

Therefore, set $\pi_\xi=\lim_{T\to\infty}E\left[s_T\left(\theta_1,\xi\right)\right]$, where $\theta_1=\operatorname{plim}\hat{\theta}$ under H_1. Moreover, in the first term $\hat{V}(\xi)^{-1/2}-V(\xi)^{-1/2}=o_p\left(1\right)$ by Lemma A.3 and Slutsky's Theorem and since $\left\|\pi_\xi\right\|=O\left(1\right)$ by Assumption 4(ii), the first term on the right-hand side of (A-16) is $o_p\left(1\right)$. Therefore,

$$\left\|\hat{V}(\xi)^{-1/2}s_T(\hat{\theta},\xi)-V(\xi)^{-1/2}\pi_\xi\right\|=o_p\left(1\right).$$

Now by Assumption 7 and Lemma 3.1, $V(\xi)^{-1/2}\pi_\xi\neq 0$ for every $\xi\in\Xi-\left(B^*\cup B_0\right)$. ∎

Proof of Theorem 3.1. The result under H_0 follows from Lemmas 3.2 and A.4. Under H_1, it follows from Lemma A.5 that $\operatorname{plim}_{T\to\infty}S_B/T=\pi_{\xi 0}'V(\xi)^{-1}\pi_{\xi 0}=\rho\left(\xi\right)$, where $\pi_{\xi,0}=\lim_{T\to\infty}E\left[s_T\left(\theta_0,\xi\right)\right]=0$ only on a set B_0 of Lebesgue measure zero defined in Lemma 3.1. Therefore, $P\left[\rho\left(\xi\right)>0\right]=1$ for each $\xi\in\Xi-B_0$. ∎

Lemma A.6 *Under Assumptions 1-4 and H_0, $\sqrt{T}z_T\left(\theta_0,\xi\right)$ defined in (A-13) is tight in $\Xi=[-b,b]^K$ for any $b>0$.*

Proof. We prove this result for the case where the weight functions w_t depend on lags of infinite order in the limit. The finite-lag case is clearly subsumed in this one.

Following Newey (1991, p1163), it suffices to prove that for all $\eta\in\mathbb{R}^p$ such that $\eta'\eta=1$,
(i) $\sqrt{T}\eta' z_T(\theta_0,\xi_0)=O_p(1)$.
(ii) For each $\delta>0$ and $\varepsilon>0$, there exists $\tau>0$ such that

$$P\left(\sup_{\left\|\xi_1-\xi_2\right\|<\tau}\left|\sqrt{T}\eta'\left(z_T(\theta_0,\xi_1)-z_T(\theta_0,\xi_2)\right)\right|\geq\varepsilon\right)\leq\delta$$

for all $T\geq T_0$, where $T_0<\infty$.

Condition (i) follows from Lemma 3.2. From (A-13) and the Markov inequality, it is sufficient for condition (ii) to hold that

$$\begin{aligned}
&E\left(\sup_{\|\xi_1-\xi_2\|<\tau}\left|\sqrt{T}\eta'\left(z_T(\theta_0,\xi_1)-z_T(\theta_0,\xi_2)\right)\right|\right) \\
&\le E\left(\sup_{\|\xi_1-\xi_2\|<\tau}\left|\sqrt{T}\eta'\left(s_T^{\infty}(\theta_0,\xi_1)-s_T^{\infty}(\theta_0,\xi_2)\right)\right|\right) \\
&\qquad + E\left(\sup_{\|\xi_1-\xi_2\|<\tau}\left|\eta'\left(\left[Q_T^{\infty}(\xi_1)-Q_T^{\infty}(\xi_2)\right]M_T^{-1}\sqrt{T}d_T(\theta_0)\right)\right|\right) \\
&\le \delta\varepsilon. && \text{(A-17)}
\end{aligned}$$

It further suffices to show that the two terms of the majorant of (A-17) are each $O(\tau)$ as $\tau \to 0$.

Note that

$$|w_t^{\infty}(\xi_1)-w_t^{\infty}(\xi_2)| \le C\tau \qquad \text{(A-18)}$$

with probability 1 for each t, by Lemma A.2. Therefore,

$$\begin{aligned}
&E\left(\sup_{\|\xi_1-\xi_2\|<\tau}\sqrt{T}\left|\eta'\left(s_T^{\infty}(\theta_0,\xi_1)-s_T^{\infty}(\theta_0,\xi_2)\right)\right|\right) \\
&= E\left(\sup_{\|\xi_1-\xi_2\|<\tau}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta' d_t(\theta_0)\left(w_t^{\infty}(\xi_1)-w_t^{\infty}(\xi_2)\right)\right|\right) \\
&= C\tau E\left(\sup_{\|\xi_1-\xi_2\|<\tau}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta' d_t(\theta_0)\left(\frac{w_t^{\infty}(\xi_1)-w_t^{\infty}(\xi_2)}{C\tau}\right)\right|\right) \\
&= O(\tau). && \text{(A-19)}
\end{aligned}$$

The final equality in (A-19) follows since $d_t(\theta_0)\left(w_t^{\infty}(\xi_1)-w_t^{\infty}(\xi_2)\right)$ is a martingale difference sequence for any $(\xi_1,\xi_2) \in \Xi \times \Xi$, and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta' d_t(\theta_0)\left(\frac{w_t^{\infty}(\xi_1)-w_t^{\infty}(\xi_2)}{C\tau}\right) \xrightarrow{d} N(0,v(\xi_1,\xi_2))$$

where $v(\xi_1,\xi_2) \le \eta'\Sigma\eta$ in view of (A-18). Since the supremum specified in the penultimate member of (A-19) over the compact set $\Xi \times \Xi$ is at a point of the set, this random variable is integrable in the limit.

Similarly

$$\begin{aligned}
&E\left(\sup_{\|\xi_1-\xi_2\|<\tau}\left|\eta'\left(\left[Q_T^{\infty}(\xi_1)-Q_T^{\infty}(\xi_2)\right]M_T^{-1}\sqrt{T}d_T(\theta_0)\right)\right|\right) \\
&= E\left(\sup_{\|\xi_1-\xi_2\|<\tau}\left|\frac{1}{T}\sum_{t=1}^{T}E\left[-\left(w_t^{\infty}(\xi_1)-w_t^{\infty}(\xi_2)\right)\eta'\left.\frac{\partial d_t(\theta)}{\partial\theta'}\right|_{\theta=\theta_0}\right]M_T^{-1}\sqrt{T}d_T(\theta_0)\right|\right) \\
&= C\tau E\left(\sup_{\|\xi_1-\xi_2\|<\tau}\left|\eta' M_T^{*}(\xi_1,\xi_2)M_T^{-1}\sqrt{T}\eta' d_T(\theta_0)\right|\right) \\
&= O(\tau) && \text{(A-20)}
\end{aligned}$$

where

$$M_T^{*}(\xi_1,\xi_2) = \frac{1}{T}\sum_{t=1}^{T}E\left[-\left(\frac{w_t^{\infty}(\xi_1)-w_t^{\infty}(\xi_2)}{C\tau}\right)\left.\frac{\partial d_t(\theta)}{\partial\theta'}\right|_{\theta=\theta_0}\right]$$

is a finite matrix, since the matrices $E\left[-\left.\frac{\partial d_t(\theta)}{\partial\theta'}\right|_{\theta=\theta_0}\right]$ are positive semi-definite and finite and the scalar weights in no case exceed 1 in absolute value. It follows from Lemma 3.2 that the expectation in the penultimate member of (A-20) is finite. ▮

Proof of Theorem 3.2. The result follows from Lemmas 3.2, A.6 and A.4. ▮

Proof of Theorem 3.3. Under H_0, since Ξ is a compact set it follows from from Theorem 3.1 that $\widehat{S}_T = O_p(1)$ and $S_{T0} = O_p\left(1\right)$. Therefore, for any $\gamma > 0$, $\rho \in \left(0,1\right)$, $P\left[\widehat{S}_T - S_{T0} > \gamma T^{\rho}\right] \to 0$ as $T \to \infty$. Thus, under H_0, $P(\tilde{S}_T = S_{T0}) \to 1$ as $T \to \infty$ and hence $\widetilde{S}_B \xrightarrow{d} \chi^2(p)$. Under H_1, the asymptotic distribution follows from Theorem 3.1.Note that $T^{-1}(\widehat{S}_T - S_{T0}) \to G(q) - q(\xi_0)$ whereas $\gamma T^{\rho-1} \to 0$ and hence the limiting choice of statistic depends on the sign of the difference. ▮

		Bierens Test (1df)			Score-based Test (3 df)		
Test Type	T	Poly-1	Poly-3	3 lags	Poly-1	Poly-3	3 lags
Sup.	200	4.45	4.56	4.29	5.07	4.91	5.33
		0.20	*0.12*	*0.22*	*0.12*	*0.03*	*0.26*
		1.72	1.39	1.36	1.45	1.34	1.62
	500	4.94	5.01	4.57	4.84	4.86	4.85
		0.18	*0*	*0*	*0.03*	*0*	*0*
		1.33	0.88	0.81	1.08	0.69	0.89
	1000	4.84	4.77	4.61	4.85	4.93	4.80
		0.04	*0*	*0.03*	*0*	*0.01*	*0*
		1.16	0.80	1.16	0.78	1.02	0.93
ICM-B	200	4.33	4.37	4.32	5.04	4.99	5.32
		0.11	*0.17*	*0.22*	*0.041*	*0.15*	*0.26*
		1.54	2.12	1.69	1.13	2.01	2.05
	500	4.83	5.08	4.70	4.82	4.88	4.94
		0.07	*0.07*	*0.07*	*0.01*	*0.02*	*0.07*
		1.18	1.62	1.47	1.02	1.19	1.72
	1000	4.82	4.79	4.62	4.85	4.92	4.80
		0.02	*0*	*0.04*	*0*	*0.02*	*0*
		1.16	0.92	1.47	0.69	1.49	0.97

Table 1: Rejection frequencies (%) of AR(1) Null Hypothesis. (See text for full details)

Test Type		T	$\tilde{B}$	$\tilde{S}$	$\tilde{S}_{\phi_0}$	$\tilde{S}_{\phi_1}$	$\hat{S}_{\sigma^2}$
Sup.	Poly-1	200	76.3 [0]	59.6 [0]	76.1 [0]	26.5 [0]	9.7 [0]
		500	98.8[0]	95.4 [0]	98.8 [0]	62.1 [0]	15.5 [0]
		1000	100 [0]	100 [0]	100 [0]	90.7 [0]	28.4 [1]
	Poly-3	200	58.2 [4]	41.7 [3]	58.9 [7]	18.1 [0]	9.4 [0]
		500	93.8 [20]	85.1 [10]	94.1 [22]	47.2 [4]	17.0 [0]
		1000	99.9 [61]	99.4 [39]	99.9 [63]	82.5 [20]	30.9 [0]
	3 lags	200	80.0 [2]	64.2 [3]	80.6 [3]	33.9 [2]	10.3 [2]
		500	99.2 [6]	97.3 [7]	99.3 [9]	74.8 [6]	15.2 [1]
		1000	100 [23]	100 [22]	100 [26]	97.0 [17]	27.3 [0]
ICM-B	Poly-1	200	76.4 [0]	59.5 [0]	76.2 [0]	26.4 [0]	9.6 [1]
		500	98.8 [0]	95.4 [0]	98.8 [0]	62.1 [0]	15.2 [0]
		1000	100 [0]	100 [0]	100 [0]	90.7 [0]	40.5 [0]
	Poly-3	200	58.1 [4]	44.3 [7]	58.3 [6]	17.7 [0]	9.5 [1]
		500	93.7 [21]	87.2 [22]	94.2 [24]	46.4 [3]	16.9 [0]
		1000	100 [69]	99.6 [67]	100 [70]	83.0 [25]	31.0 [0]
	3 lags	200	80.0 [2]	65.5 [6]	80.9 [4]	33.2 [1]	9.3 [1]
		500	99.2 [8]	97.39 [17]	99.2 [11]	74.7 [6]	15.2 [0]
		1000	100 [29]	100 [45	100 [31]	97.4 [20]	27.4 [2]

Table 2: Rejection frequencies (%) : AR(1) model fitted to ESTAR series

Test	$AR(1)$	$AR(2)$	$AR(3)$	$AR(4)$
Ljung-Box (12)	82.6	6.5	5.8	5.1
McLeod-Li (12)	8.8	8.2	7.8	7.6
Breusch-Godfrey LM (4)	92.3	3.4	1.5	7.6
Engle ARCH LM (4)	10.1	10.0	10.0	9.2
Ramsey RESET (Sq)	10.4	48.2	46.8	44.7
$\tilde{B}$	99.2	57.8	65.9	67.0
$\tilde{S}$	97.5	70.7	66.7	63.9
$\tilde{S}_{\phi_0}$	95.0	58.0	65.6	66.3
$\tilde{S}_{\phi_1}$	74.8	59.9	59.0	58.8
$\tilde{S}_{\phi_2}$	-	84.4	84.6	84.2
$\tilde{S}_{\phi_3}$	-	-	5.8	5.6
$\tilde{S}_{\phi_4}$	-	-	-	6.4
$\tilde{S}_{\sigma^2}$	15.2	6.1	6.3	6.3

Table 3: Rejection frequencies, AR(p) models fitted to ESTAR, ICM-B test, 3 lags, $T = 500$

T	Model	$\tilde{B}$	$\tilde{S}$	$\tilde{S}_{\phi_0}$	$\tilde{S}_{\phi_1}$	$\tilde{S}_{\phi_2}$	$\tilde{S}_{\theta_1}$	$\tilde{S}_{\sigma^2}$	$\tilde{S}_{\alpha}$	$\hat{S}_{\beta}$
200	ARMA	4.95	6.04	3.50	4.50	-	4.6	6.92	-	-
	AR2	4.01	6.50	2.16	4.26	3.96	-	6.86	-	-
	GARCH1	4.48	5.02	5.62	-	-	-	7.36	3.32	2.92
	AR-GARCH	1.71	3.64	2.49	2.68	-	-	4.93	1.51	0.97
500	ARMA	4.98	5.61	4.52	4.68	-	4.64	6.24	-	-
	AR2	4.88	5.36	3.76	4.59	4.77	-	6.14	-	-
	GARCH1	4.80	3.92	5.48	-	-	-	6.23	3.76	3.86
	AR-GARCH	3.21	4.75	5.42	3.65	-	-	6.65	1.84	0.51

Table 4: Rejection frequencies (%) under the null hypothesis: IBM-B test with 3 lags

Simulated Model	$\tilde{B}$	$\tilde{S}$	$\tilde{S}_{\phi_0}$	$\tilde{S}_{\phi_1}$	$\tilde{S}_{\sigma^2}$
AR2	99.0	98.7	98.5	15.6	6.8
ARMA	58.1	58.6	57.4	10.0	6.6
SETAR	100	100	100	100	6.3
SGN	81.6	99.9	76.	44.5	15.0
BILIN	13.1	99.8	6.9	21.8	99.0
NLMA	35.2	58.2	33.5	75.9	6.9
MARKOV-SW	100	100	100	98.8	83.8
ARCH	6.5	34.3	9.3	8.2	34.8
GARCH	5.2	20.7	6.9	6.2	26.8

Table 5: Rejection frequencies (%) : Sup test with 3 lags in AR(1) estimation, $T = 500$

Simulated Model	$\hat{B}$	$\tilde{S}$	$\tilde{S}_{\phi_0}$	$\tilde{S}_{\gamma}$	$\tilde{S}_{\alpha}$	$\tilde{S}_{\beta}$
EGARCH	7.3	84.2	6.8	96.4	65.8	55.8
GJR	12.7	46.9	6.8	49.0	62.4	19.9

Table 6: Rejection frequencies (%) in tests of the GARCH(1,1). Sup test, 3 lags, $T = 500$

Test Type		$\tilde{B}$	$\tilde{S}$	$\tilde{S}_{\phi_0}$	$\tilde{S}_{\phi_1}$	$\tilde{S}_{\sigma^2}$
Sup	Poly-1	59.1	52.7	60.6	8.7	7.4
	Poly-3	46.5	39.7	53.6	7.3	7.5
	Poly-6	92.9	78.3	94.8	9.1	6.5
	3 Lags	72.4	68.2	75.9	9.8	8.8
	6 Lags	99.3	98.4	99.5	17.1	8.0
ICM-B	Poly-1	55.3	52.5	56.7	8.6	7.2
	Poly-3	52.0	50.8	58.2	7.3	7.5
	Poly-6	96.3	93.3	97.1	8.2	6.3
	3 Lags	71.6	69.6	74.4	9.7	7.8
	6 Lags	99.2	98.5	99.4	17.9	7.1

Table 7: Rejection frequencies (%), Sup tests of the AR(1) against ARMA(1,1) alternative, $T = 500$