

Department of Econometrics and Business Statistics

<http://business.monash.edu/econometrics-and-business-statistics/research/publications>

Hypothesis testing based on a vector of statistics

Maxwell L. King, Xibin Zhang and Muhammad Akram

December 2019

Working Paper 30/19

Hypothesis testing based on a vector of statistics

Maxwell L. King, Xibin Zhang¹

Department of Econometrics and Business Statistics, Monash University, Australia

Muhammad Akram

Mary MacKillop Institute for Health Research, Australian Catholic University

August 31, 2017

Revised: July 1, 2018 and May 6, 2019

This version: November 14, 2019

Abstract: This paper presents a new approach to hypothesis testing based on a vector of statistics. It involves simulating the statistics under the null hypothesis and then estimating the joint density of the statistics. This allows the p -value of the smallest acceptance region test to be estimated. We prove this p -value is a consistent estimate under some regularity conditions. The small-sample properties of the proposed procedure are investigated in the context of testing for autocorrelation, testing for normality, and testing for model misspecification through the information matrix. We find that our testing procedure has appropriate sizes and good powers.

JEL classification: C01, C12, C14.

Keywords: Bootstrap, Cross-market prediction, Information matrix test, Markov chain Monte Carlo, Multivariate kernel density, p -value

¹Address: 900 Dandenong Road, Caulfield East, Victoria 3145, Australia. Telephone: +61 3 99032130. Fax: +61 3 99032007. Email: xibin.zhang@monash.edu.

1 Introduction

Statistical hypothesis testing is an extremely important technique in the practice of econometrics, particularly with respect to diagnostic checking of model specification. This is how econometricians are best able to combat the severe problem of uncertainty in model specification. Such testing procedures need to be as accurate as possible due to constraints on data availability. Fortunately, advances in computer power and simulation based methods have allowed greater scope in the design of high quality tests. The purpose of any test is to accurately control the probability of wrongly rejecting the null hypothesis (known as the size of the test), while at the same time ensuring a high probability of correctly rejecting the null hypothesis (known as the power of the test).

There are a number of popular tests whose construction involves a desire to determine whether a $d \times 1$ vector of statistics denoted by $\mathbf{t}(\mathbf{y})$, where \mathbf{y} is a $T \times 1$ vector of observations, is significantly different from zero. Examples include the Lagrange multiplier (LM) test (see [Godfrey, 1991](#)) where $\mathbf{t}(\mathbf{y})$ is a vector of Lagrange multipliers, the Wald test (see [Engle, 1984](#)) where $\mathbf{t}(\mathbf{y})$ is a vector of parameter estimates minus their null hypothesis values, the Durbin-Wu-Hausman test (see [Davidson and MacKinnon, 1993](#), Section 7.9) in which $\mathbf{t}(\mathbf{y})$ is a vector of differences between two types of parameter estimates, the information matrix test ([White, 1982](#)) where $\mathbf{t}(\mathbf{y})$ is a vector of differences between the corresponding elements of the negative Hessian matrix and the outer product of the score vector, and conditional moment tests (see [Pagan and Vella, 1989](#)) where $\mathbf{t}(\mathbf{y})$ is a set of sample moments. The general method of test construction then involves determining the asymptotic covariance matrix V of $\mathbf{t}(\mathbf{y})$ and using a consistent estimate of V , denoted as \hat{V} , to obtain the quadratic form

$$\mathbf{t}(\mathbf{y})' \hat{V}^{-1} \mathbf{t}(\mathbf{y}). \tag{1}$$

Under suitable regularity conditions, the latter can be shown to have a χ^2 asymptotic distribution when the null hypothesis holds, which can be used to test whether $\mathbf{t}(\mathbf{y})$ is significantly different from zero.

Unfortunately, these tests can have poor size and power properties, particularly in small samples, because of \hat{V} being a poor estimate of the true covariance matrix of $t(\mathbf{y})$. Research that have found such outcomes include papers by [Gregory and Veall \(1985, 1986, 1987\)](#), [Lafontaine and White \(1986\)](#), [Mantel \(1987\)](#), [Taylor \(1987\)](#), [Moulton and Randolph \(1989\)](#), [Orme \(1990\)](#), [Chesher and Spady \(1991\)](#), [Davidson and MacKinnon \(1992\)](#) and [Dhaene and Hoorelbeke \(2004\)](#). In the case of testing non-linear restrictions in the parameters of a linear model using the Wald test, [Breusch and Schmidt \(1988\)](#) were able to demonstrate the possibility of obtaining any value for the Wald statistic by rewriting the restrictions. Another problem that has received less attention involves the mean of $t(\mathbf{y})$ being different from its asymptotic mean. [Goh and King \(1999\)](#) suggested this is a possible cause of local biasedness (power lower than size) in the Wald test. A third issue is that a test of the form of (1) assumes $t(\mathbf{y})$ is multivariate normal. While this is likely to be the case asymptotically, it cannot be guaranteed in small samples. This list of problems does raise the question of whether (1) is the most appropriate test statistic in finite samples.

There are similarities with this particular test construction problem and multiple testing of a common null hypothesis using the same data also known as induced testing (or combined testing). A very large and growing literature on diagnostic testing of all kinds of econometric models (see, for example, [Breusch and Pagan, 1980](#); [Engle, 1984](#); [Tauchen, 1985](#); [Wooldridge, 2001](#); [Godfrey, 2009](#)) allows researchers to check the adequacy of a chosen model by applying a range of diagnostic tests, each of which is designed to detect a particular form of model inadequacy. A major problem is how best to control the overall probability of rejecting the model when it is true. For example, five statistically independent tests applied at the 5% level will result in a 22.6% chance of at least one rejection when the null hypothesis model is true. Of course, it is unlikely that five diagnostic tests applied to the same model will be mutually independent, so in actual fact, this probability could be higher or lower than 22.6%. The core issue is how we should conduct these tests in order to control the overall probability of rejecting the model when it is true with an eye to having good power to reject the model when it is not true. For this type of

application, $t(\mathbf{y})$ denotes the vector of test statistics.

An early solution to this problem involves the use of the Bonferroni inequality to set an upper bound on the overall probability of rejecting the model. Unfortunately, this approach is known to be conservative and therefore to have reduced power, particularly when some of the individual tests are highly correlated. There have been a number of attempts to improve this approach (see, for example, [Holm, 1979](#); [Simes, 1986](#)), but it is still regarded as a conservative approach to the problem. We have also seen the development of joint LM tests which test for multiple forms of misspecification (see, for example, [Bera and Jarque, 1982](#); [Baltagi and Li, 1995](#); [Baltagi, Bresson and Pirotte, 2006](#); [Baltagi, Jung and Song, 2010](#)).

[Dufour and Khalaf \(2001\)](#) and [Dufour \(2006\)](#)² developed a framework to address this problem based on Monte Carlo testing introduced by [Dwass \(1957\)](#) and [Barnard \(1963\)](#). It requires being able to calculate a p -value (or approximate p -value) for each individual test statistic $t_i(\mathbf{y})$ and then combining the d p -values either by calculating the minimum p -value, the product of p -values or a weighted product of p -values (or some other variation of these approaches) to produce a single value that can be used as the test statistic in a Monte Carlo test. The latter involves a Monte Carlo simulation to calculate an empirical p -value of this new test statistic. When it can be shown that all the individual tests that give rise to the d p -values are similar tests (i.e., their distributions under the null hypothesis do not depend on nuisance parameters), then this approach leads to an exact α -level test provided $\alpha(n + 1)$ is an integer where α is the desired significance level and n is the number of Monte Carlo replications.

In the case where the null distribution of $t(\mathbf{y})$ depends on nuisance parameters, [Dufour \(2006\)](#) proposed the maximized Monte Carlo (MMC) test which involves maximizing the simulated p -value with respect to the nuisance parameters. This results in an exact test in the sense that the test's size, which is a function of the nuisance parameters, is always less than or equal to α . Unfortunately, because the size of this test is likely to be less than α at the true (unknown)

²Also see [Dufour, Farhat, Gardiol and Khalaf \(1998\)](#); [Dufour and Khalaf \(2002\)](#); [Dufour, Khalaf and Beaulieu \(2003, 2010\)](#); [Dufour, Khalaf, Bernard and Genest \(2004b\)](#); [Dufour, Farhat and Khalaf \(2004a\)](#); [Beaulieu, Dufour and Khalaf \(2007, 2009\)](#); [Bernard, Idoudi, Khalaf and Yélou \(2007\)](#); [Dufour, Khalaf and Voia \(2015\)](#).

value of the nuisance parameters, it is likely to be a conservative test with some loss of power. In view of this, Dufour also suggested the Consistent Set Estimate MMC test, which involves restricting the maximization search to a confidence set for the nuisance parameters under the null hypothesis. A further suggestion involves obtaining a consistent estimate of the nuisance parameters and using these estimates in the simulations to apply the Monte Carlo test. Both alternatives have asymptotic justifications.

Bootstrap solutions have been outlined by [Westfall and Young \(1993\)](#), [Godfrey \(2005\)](#) and [Godfrey \(2009, Section 4.3\)](#). Godfrey's procedure involves first calculating p -values for each of the individual tests, preferably via a series of first-stage bootstrap sampling. These p -values are then combined via the minimum p -values approach and a second-stage of bootstrap sampling is used to calculate an approximate overall p -value that is asymptotically justified.

This paper proposes an alternative approach to hypothesis testing based on a vector of statistics, $\mathbf{t}(\mathbf{y}): R^T \rightarrow R^d$ ($T > d$). It assumes that each of the elements of $\mathbf{t}(\mathbf{y})$ has been chosen because individually they have good power to detect a particular deviation from the null hypothesis. It is further assumed that collectively $\mathbf{t}(\mathbf{y})$ provides a good summary of the evidence contained in \mathbf{y} that might point to the null hypothesis being false. The approach involves asking the question based on the observed value of $\mathbf{t}(\mathbf{y})$, do we think the null hypothesis is true? If we know the joint density function for $\mathbf{t}(\mathbf{y})$ under the null hypothesis denoted as $f(\mathbf{t})$, then following [Hyndman \(1996\)](#) who discusses the calculation of highest density regions for multivariate densities, we can calculate the p -value for the observed value of $\mathbf{t}(\mathbf{y})$. This is equivalent to using the highest density region of $f(\mathbf{t})$ of probability $(1 - \alpha)$ to determine the acceptance region of the test. Typically we do not know this joint density function, and even if we did, it is possible that $f(\mathbf{t})$ will depend on unknown nuisance parameter values. Our approach is either to simulate independent values of $\mathbf{t}(\mathbf{y})$ under the null hypothesis using consistent estimates of any nuisance parameters that $f(\mathbf{t})$ depends on or to use bootstrap sampling to calculate a random sample of $\mathbf{t}(\mathbf{y})$ values, and then use a multivariate kernel density estimator to estimate the density.

The contribution this paper makes is to propose a way to (approximately) control the prob-

ability of falsely rejecting the null hypothesis based on a vector of statistics. This is done by approximating the smallest acceptance region (SAR) test. We prove that under some regularity conditions, the estimated p -value of our proposed test procedure is a consistent estimate of the true p -value of the SAR test, thus proving our test is an asymptotic SAR test as the sample size and numbers of simulations all go to infinity. We are also able to infer that our test statistic based on the estimated joint density of $\mathbf{t}(\mathbf{y})$ is asymptotically pivotal. When testing a subvector of coefficient estimates in the classical linear regression model, $f(\mathbf{t})$ is known and we show a version of our test is equivalent to the F test. We also observe that our approach has the potential to eliminate one source of biasedness in the Wald test. We show that one version of our test can be regarded as a Monte Carlo test as outlined by [Dufour \(2006\)](#) and therefore can be an exact test when $f(\mathbf{t})$ does not depend on unknown nuisance parameters. Also in this case, we observe that the SAR test can be approximated by our procedure with arbitrary accuracy by using a sufficiently large number of iterations in the required simulations. When $f(\mathbf{t})$ does depend on nuisance parameters, [Dufour's \(2006\)](#) MMC procedure can be applied to a version of our test to yield an exact, but potentially conservative test. We also investigate the small-sample size and power properties of the new test through simulation studies. One revealing study involves applying the procedure to the information matrix test. The new procedure's estimated sizes are typically not significantly different from their nominal sizes, and its powers are almost always higher than, and in some cases more than double that of the best of six existing tests.

The rest of this paper is organized as follows. Section [2](#) explains the new testing procedure when $f(\mathbf{t})$ is independent of nuisance parameters and discusses its properties. In Section [3](#), we present the testing procedure when $f(\mathbf{t})$ depends on nuisance parameters. In Section [4](#), we examine the performance of the new testing procedure through Monte Carlo simulations by comparing its size and power with that of a well known existing test and two-sided versions of [Dufour's \(2006\)](#) minimum p -value and product of p -value tests. Section [5](#) briefly describes the information matrix test and its limitations. We present a simulation study of the new testing procedure applied to the information matrix test in Section [6](#). An empirical application of the

proposed test is conducted through the information matrix test in Section 7. Section 8 concludes the paper.

2 The test procedure for statistics independent of nuisance parameters

We shall begin by first describing the main ideas behind our new testing procedure. Assume that we wish to test the null hypothesis that the $T \times 1$ vector of observations \mathbf{y} has a particular fully parametric or semi-parametric³ data generating process (DGP) using d statistics denoted as t_i , for $i = 1, 2, \dots, d$. Let $\mathbf{t} = (t_1, t_2, \dots, t_d)'$ represent the $d \times 1$ vector of statistics called the test vector hereafter, and let $\hat{\mathbf{t}}$ denote the calculated value of \mathbf{t} using the available data. We also assume that each of the component tests is based on accepting the null hypothesis for the most likely observed values of t_i purely based on its distribution under the null. This rules out one-sided testing⁴ and results in a two-sided test when the null distribution is unimodal. It is a natural consequence of tests of the form of (1) which treats $\mathbf{t}(\mathbf{y})$ and $-\mathbf{t}(\mathbf{y})$ equally so these tests are two-sided with respect to $\mathbf{t}(\mathbf{y})$. Let the joint density function of \mathbf{t} under the null hypothesis be denoted by $f(\mathbf{t})$. For the moment, we assume that $f(\mathbf{t})$ does not depend on nuisance parameters.

Essentially, we wish to ask the question, is the observed value of $\hat{\mathbf{t}}$ consistent with the null hypothesis being true? The p -value is a useful tool for answering this question. It is defined as the probability under the null hypothesis of finding a value of the test vector as extreme as or more extreme than the value we have found from the data, namely $\hat{\mathbf{t}}$. Thus, if we have $f(\mathbf{t})$, the p -value of the test vector is the probability of obtaining a value of \mathbf{t} such that $f(\mathbf{t}) < f(\hat{\mathbf{t}})$ holds. Once calculated, the p -value can be used to conduct the test at any level of significance. For example, at the 5% significance level, if the p -value is less than 0.05 then the null hypothesis is rejected. Otherwise, it cannot be rejected. This test can also be viewed as rejecting the null hypothesis for significantly small values of $f(\hat{\mathbf{t}})$. The resultant acceptance region is optimal in the sense that by

³Semi-parametric DGPs are restricted to those that allow a bootstrap to be applied that is consistent (see Horowitz, 2001).

⁴Some preliminary results for one-sided component tests are reported in King et al. (2009).

its construction, it is the smallest $(1 - \alpha)$ acceptance region in the d -dimensional sample space of \mathbf{t} . This is because the acceptance region is the highest density region of probability $(1 - \alpha)$ based on $f(\mathbf{t})$. If we believe $\mathbf{t}(\mathbf{y})$ provides a good summary of the evidence contained in \mathbf{y} that might point to the null hypothesis being false then this is a desirable property to have. We call a test with this property a smallest acceptance region (SAR) test.

To better appreciate the desirability of this approach, consider the classical regression

$$\mathbf{y} = X_1\beta_1 + X_2\beta_2 + \mathbf{u}, \quad \mathbf{u} \sim N(\mathbf{0}, \sigma^2 I), \quad (2)$$

where \mathbf{y} is $T \times 1$, X_1 is $T \times j$, X_2 is $T \times d$, β_1 is $j \times 1$, β_2 is $d \times 1$ and \mathbf{u} is $T \times 1$. Assume we wish to test $H_0 : \beta_2 = 0$ based on the ordinary least squares (OLS) estimate, $\hat{\beta}_2$. It can be shown that

$$\hat{\beta}_2 \sim N\left(\beta_2, \sigma^2 (X_2' M_1 X_2)^{-1}\right),$$

where $M_1 = I - X_1(X_1' X_1)^{-1} X_1'$. So if we wish to test H_0 based on $\mathbf{t} = \hat{\beta}_2$, $f(\mathbf{t})$ is known up to a scalar parameter, σ^2 . In Appendix 1, we show that when σ^2 is replaced by its unbiased estimate,

$$\hat{\sigma}^2 = \mathbf{z}' \mathbf{z} / (T - k),$$

where \mathbf{z} is the OLS residual vector and $k = j + d$, our proposed test is equivalent to the well known F test. As [Seber \(1980\)](#) notes, the F test has a number of optimal properties including being uniformly most powerful invariant under a range of invariant transformations.

Also note that the distribution of \mathbf{t} (conditional on σ^2), only involves a change in mean as we move from the null hypothesis being true to the alternative ($\beta_2 \neq 0$) being true. For testing problems that are based on statistics whose distributions are free of nuisance parameters and only involve a change in mean as we move from the null hypothesis to the alternative, a SAR test is also an unbiased test with power always greater than size. The Wald test is known to be locally biased (see [Peers, 1971](#); [Hayakawa, 1975](#); [King and Goh, 2002](#)). [Goh and King \(1999\)](#) postulate that a possible cause is bias in the estimates of the parameters under test. For a test of the form of (1), if elements of $\mathbf{t}(\mathbf{y})$ are biased under the null with non-zero means, then this will be reflected

directly in $f(\mathbf{t})$. Calculating p -values using $f(\mathbf{t})$ rather than the null distribution of (1) can be expected to eliminate this possible source of local biasedness.

Theorem 1: If we know that $\mathbf{t} \sim N(\mathbf{a}, W)$ under the null hypothesis, where \mathbf{a} is a known $d \times 1$ vector and W is a known $d \times d$ positive definite matrix, then our SAR test based on rejecting the null hypothesis for small values of $f(\hat{\mathbf{t}})$ is equivalent to the χ^2 test based on rejecting the null hypothesis for large values of $(\hat{\mathbf{t}} - \mathbf{a})'W^{-1}(\hat{\mathbf{t}} - \mathbf{a})$.

Proof: The proof of Theorem 1 is given in Appendix 2.

Corollary 1: If we know that $\mathbf{t} \sim N(\mathbf{0}, V)$ under the null hypothesis, where V is a known $d \times d$ positive definite matrix, then our SAR test is equivalent to the χ^2 test based on rejecting the null hypothesis for large values of $\hat{\mathbf{t}}'V^{-1}\hat{\mathbf{t}}$.

The significance of these results are as follows. The conditions of Corollary 1 typically only happen when the sample size, T , approaches infinity implying that these χ^2 tests are asymptotically SAR tests. Suppose that $\hat{\mathbf{t}} \sim N(\mathbf{a}, W)$ where $\mathbf{a} \rightarrow \mathbf{0}$ and $W \rightarrow V$ as $T \rightarrow \infty$ as is the case for standard χ^2 tests. These tests are based on rejecting the null for large values of $\hat{\mathbf{t}}'V^{-1}\hat{\mathbf{t}}$, while Theorem 1 tells us that our proposed SAR test is equivalent to rejecting for large values of $(\hat{\mathbf{t}} - \mathbf{a})'W^{-1}(\hat{\mathbf{t}} - \mathbf{a})$ which in these circumstances would be the preferred test. The standard χ^2 tests are based on a form of the test statistic that may only be appropriate in the limit. Our test, in this case, is equivalent to using the test statistic that is appropriate for the small sample distribution of \mathbf{t} under the null.

The assumption that \mathbf{t} is always multivariate normal allowed us to make this observation. This frequently may not be the case in small samples. If we know $f(\mathbf{t})$, our approach provides a SAR test in the case of non-normality. We would argue that the potential for \mathbf{t} to be non-normal, strengthens our argument for a test based on $f(\hat{\mathbf{t}})$.

Typically the density $f(\mathbf{t})$ is unknown. We can estimate it in the case of a fully parameterized DGP by applying a multivariate kernel density estimator to a sample of independent drawings from $f(\mathbf{t})$ which can be obtained by repeatedly simulating the DGP under the null hypothesis and then calculating \mathbf{t} for each simulated data set. Because $f(\mathbf{t})$ does not depend on unknown

parameters, we can set unknown parameters in the DGP to any value we wish in order to conduct these simulations. Let $\{t_1, t_2, \dots, t_m\}$ denote such a sample. The general form of the kernel density estimator of t is given by

$$\hat{f}_{m,H}(t) = \frac{1}{m} \sum_{i=1}^m |H|^{-1/2} K(H^{-1/2}(t - t_i)), \quad (3)$$

where $K(\cdot)$ is a kernel function, and H is a positive definite matrix of bandwidths known as the bandwidth matrix (see, for example, [Wand and Jones, 1995](#); [Scott, 2015](#)). Note that if the conditions hold for which $\hat{f}_{m,H}(t)$ is uniformly consistent on a bounded set in which $f(t) > 0$ (see Section 3.3 below), then for those t values of interest, $f(t)$ can be approximated by $\hat{f}_{m,H}(t)$ with arbitrary accuracy by choosing m to be sufficiently large.

In the case of a semi-parametric DGP, bootstrap methods can be used to obtain a random sample of drawings from a d -dimensional distribution that is close to $f(t)$. There is a large and growing literature on bootstrap sampling. See for example, the books by [Hall \(1992\)](#) and [Godfrey \(2009\)](#) and review articles written for econometricians by [Hall \(1994\)](#), [Dufour and Khalaf \(1997\)](#); [Horowitz \(2001, 2018\)](#) and [MacKinnon \(2009\)](#). A multitude of different procedures have been proposed, some being specific to particular settings. If a bootstrap procedure that is consistent (see Horowitz, 2001) is chosen, then the bootstrap sample can be regarded as converging almost surely to a random sample from $f(t)$ as $T \rightarrow \infty$. If we denote this sample as $\{t_1, t_2, \dots, t_m\}$, then we can use (3) to estimate $f(t)$.

There are two ways in which the new testing procedure can be implemented in practice. The first, which we call the double simulation method (DSM), involves two separate rounds of simulation as follows:

- (i) Based on the data under test, calculate \hat{t} .
- (ii) Compute the random sample t_1, t_2, \dots, t_m either by simulating the null hypothesis DGP m times to calculate m independent values of t in the fully parameterized case⁵, or by drawing

⁵If, as is likely, the DGP in this case has unknown parameter values, these can be set to any convenient value (such as zero or one) because $f(t)$ does not depend on their values. We get exactly the same sample, t_1, t_2, \dots, t_m , for different choices of parameter values.

a bootstrap sample under the null hypothesis in the case of the semi-parametric DGP.

- (iii) Use the sample generated in (ii) to estimate the joint density $f(\mathbf{t})$ by $\hat{f}_{m,H}(\mathbf{t})$ via (3).
- (iv) Repeat (ii) to generate a second sample of n values of \mathbf{t} denoted as $\mathbf{t}^{(1)}, \mathbf{t}^{(2)}, \dots, \mathbf{t}^{(n)}$, which are independent of those originally calculated in step (ii).
- (v) Use this second sample to calculate $\hat{f}_{m,H}(\mathbf{t}^{(i)})$, for $i = 1, 2, \dots, n$. The p -value of the joint test is estimated by the relative frequency for which

$$\hat{f}_{m,H}(\mathbf{t}^{(i)}) < \hat{f}_{m,H}(\hat{\mathbf{t}}) \quad (4)$$

holds.⁶

The second test procedure involves only one round of simulation and is therefore called the single simulation method (SSM). After completing steps (i) and (ii) above, the remaining steps are as follows:

- (iii') Use the sample generated in (ii) to estimate the joint density $f(\mathbf{t})$ at $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_m$ by the leave-one-out kernel density estimator

$$\hat{g}_{m,H}(\mathbf{t}_i) = \frac{1}{m-1} \sum_{j=1; j \neq i}^m |H|^{-1/2} K(H^{-1/2}(\mathbf{t}_i - \mathbf{t}_j)). \quad (5)$$

- (iv') The p -value of the joint test is estimated by the relative frequency for which $\hat{g}_{m,H}(\mathbf{t}_i) < \hat{f}_{m,H}(\hat{\mathbf{t}})$ holds.

If $f(\mathbf{t})$ were known, the p -value of the proposed SAR test procedure would be

$$p_0 = \Pr \{ \mathbf{t} : f(\mathbf{t}) < f(\hat{\mathbf{t}}) | f(\mathbf{t}) \}. \quad (6)$$

In our testing procedure, the p -value denoted as $\hat{p}_{T,m}$, is defined via the kernel density estimator under the probability measure implied by the true density of \mathbf{t} . This means that

$$\hat{p}_{T,m} = \Pr \{ \mathbf{t} : \hat{f}_{m,H}(\mathbf{t}) < \hat{f}_{m,H}(\hat{\mathbf{t}}) | f(\mathbf{t}) \}, \quad (7)$$

⁶See [Racine and MacKinnon \(2007\)](#) for a variation on this approach to estimating the p -value that can be advantageous when n is small.

which is estimated by the relative frequency with which (4) holds. This can be done with arbitrary accuracy by selecting n to be sufficiently large.

Equation (6) defines a low density region which is the complement of a high density region.⁷ A consequence is that the procedure can naturally handle situations in which $f(t)$ is bimodal or multimodal. In our proposed testing procedure, $\hat{p}_{T,m}$ is approximated by $\hat{p}_{T,m,n}$ that is the relative frequency of observing $\hat{f}_{m,H}(t) < \hat{f}_{m,H}(\hat{t})$ in the second-round simulation of Step (v). Because we need to estimate $f(t)$ and p_0 , the resultant test is no longer a SAR test, but we can consider it to be an approximate SAR test.

The DSM using parametric simulation can be regarded as a Monte Carlo test as it fulfills the requirements outlined by Dufour (2006). To confirm this, we need to view $\hat{f}_{m,H}(t)$ as our test statistic and observe that

$$\hat{f}_{m,H}(t^{(i)}), \text{ for } i = 1, 2, \dots, n \text{ and } \hat{f}_{m,H}(\hat{t})$$

all share the same distribution and are exchangeable by construction. Because neither the function $\hat{f}_{m,H}(\cdot)$ nor $t^{(i)}$, for $i = 1, 2, \dots, n$, depend on unknown parameter values, this shared distribution is independent of unknown parameters. In the case where there is a zero probability that any two $\hat{f}_{m,H}(t^{(i)})$ values, for $i = 1, 2, \dots, n$, will be equal, our test procedure applied at the α level of significance is an exact test provided $\alpha(n+1)$ is an integer (Dufour, 2006, Proposition 2.2). This does not require n to be particularly large although small n values can result in a small power loss as noted by Racine and MacKinnon (2006). It also applies for any value of d , no matter how large, and does not require m to be large or $\hat{f}_{m,H}(\cdot)$ to be particularly accurate. If there is a non-zero probability of two $\hat{f}_{m,H}(t^{(i)})$ values being equal, then our procedure can be amended as outlined in Dufour's (2006) Proposition 2.3 to make it exact.

There is also an issue of determining the bandwidth matrix H in the kernel density estimators (3) and (5). In this paper, we use a diagonal bandwidth matrix, $H = \text{diag}\{h_1^2, h_2^2, \dots, h_d^2\}$ and two

⁷Samworth and Wand (2010) investigated the asymptotic properties of the plug-in type bandwidth selector for kernel density estimation that aims to define a high density region. Recent studies on high density regions include Tsybakov (1997), Cadre (2006), Mason and Polonik (2009) and Cadre, Pelletier and Pudlo (2013) among others.

different methods for determining h_i . The first is the normal reference rule (NRR)

$$h_i = \hat{\sigma}_i \left\{ \frac{4}{(d+2)m} \right\}^{1/(d+4)}, \quad \text{for } i = 1, 2, \dots, d,$$

where $\hat{\sigma}_i$ is the estimated standard deviation of the i th element of \mathbf{t} . This is the optimal bandwidth estimator when \mathbf{t} is multivariate normal (see [Scott, 2015](#); [Bowman and Azzalini, 1997](#)). The second method is to use [Zhang et al.'s \(2006\)](#) Markov chain Monte Carlo (MCMC) sampling algorithm. The bandwidth matrix chosen through this algorithm tends to produce a more accurate density estimator than that chosen through the NRR. However, the MCMC bandwidth selector is far more computationally costly. A general guideline for selecting one of the two bandwidth selectors is as follows. When the required computing time is not of serious concern, one may use the MCMC bandwidth selector. Otherwise, one may use the NRR to choose bandwidths. In addition, for both methods, we use the product of d univariate Gaussian kernels as our choice of kernel function in (3) and (5).

3 The testing procedure for non-invariant test statistics

3.1 The fully parametric case

So far we have concentrated on statistics whose distribution under the null hypothesis does not depend on nuisance parameters. In the fully parametric case, there is no issue of how to simulate \mathbf{t} under the null hypothesis in order to estimate its density because we can choose any null hypothesis DGP for this purpose. Also we can approximate the SAR test with arbitrary accuracy by our choice of m and n . When the distribution of \mathbf{t} under the null hypothesis depends on the value of one or more nuisance parameters, which we denote by γ , then we have a more complex testing problem. In particular, the unknown density $f(\mathbf{t})$ now becomes a function of γ and will be denoted as $f(\mathbf{t}, \gamma)$. We are assuming that the DGP under the null hypothesis is fully parameterized in the sense that if γ is known then $f(\mathbf{t}, \gamma)$ is a single distribution for all DGPs under the null hypothesis with this value of γ .

Conditional on a value for γ , denoted as γ^* , we can apply either of the procedures outlined

in Section 2, by simulating the DGP under the null hypothesis, with γ being set to γ^* and any remaining parameters being set to convenient values because $f(\mathbf{t}, \gamma)$ does not depend on their values. The kernel density $\hat{f}_{m,H}(\mathbf{t})$ given by (3) and the leave-one-out kernel density estimator (5) now become functions of γ^* , denoted as $\hat{f}_{m,H}(\mathbf{t}, \gamma^*)$ and $\hat{g}_{m,H}(\mathbf{t}, \gamma^*)$, respectively. In both cases, the resultant p -value is dependent on γ^* and does not result in a test that effectively controls size, either exactly or approximately, unless γ^* is chosen with controlling size in mind.

In the case of the DSM, we can view $\hat{f}_{m,H}(\mathbf{t}, \gamma^*)$ as our statistic and note that there is a value of γ^* (namely γ_0 , the true value of γ under the null hypothesis) for which, conditional on γ^* ,

$$\hat{f}_{m,H}(\mathbf{t}^{(i)}, \gamma^*), \text{ for } i = 1, 2, \dots, n \text{ and } \hat{f}_{m,H}(\hat{\mathbf{t}}, \gamma^*)$$

all share the same distribution and are exchangeable. Assuming there is a zero probability of any two $\hat{f}_{m,H}(\mathbf{t}^{(i)}, \gamma^*)$ values being exactly equal, this implies that our DSM test procedure fulfils the conditions of [Dufour's \(2006\) Proposition 4.1](#) so that the MMC test is an exact α -level test provided n is chosen so that $\alpha(n+1)$ is an integer. This is true for any value of d , no matter how large.

If we denote the p -value from the DSM with $\gamma = \gamma^*$ as $\hat{p}_{T,m,n}(\gamma^*)$, the MMC test in this case involves calculating

$$\sup_{\gamma^*} \hat{p}_{T,m,n}(\gamma^*) \tag{8}$$

and using it as the test's p -value. As noted in the Introduction, the MMC test can be a conservative test if the probability of a Type I error at $\gamma = \gamma_0$ is below the desired level, α . Also, our experience with the simulated annealing algorithm which is used to solve (8), is that it can be a very computationally demanding algorithm. [Dufour's \(2006\) Consistent Set Estimate MMC test](#) mitigates these concerns to some degree although this test is only exact asymptotically.

The ideal solution is to set γ^* to the true value of γ , namely γ_0 , under the null hypothesis. Unfortunately, it is very unlikely we will know γ_0 . Our preferred approach involves what might be considered the next best option and that is to set γ^* equal to a consistent estimate of γ under the null. This involves the following variation to Step (ii) in the procedure given in Section 2:

(ii') Estimate γ using the available T observations and assuming the null hypothesis is true⁸ and denote this estimate as $\hat{\gamma}_T$. Using $\gamma = \hat{\gamma}_T$ and any convenient values of the remaining parameters in the model under the null hypothesis, simulate the model m times and calculate m independent values of \mathbf{t} denoted as $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_m$.

The remainder of the procedure is as outlined in Section 2. There can be situations in which it is difficult to determine which parameters \mathbf{t} depends on under the null hypothesis. In these cases, we can use estimated values of all the model's parameters when simulating \mathbf{t} , without any loss.

This approach involves sampling from $f(\mathbf{t}, \hat{\gamma}_T)$ in order to estimate $f(\mathbf{t}, \gamma)$ and thus can be viewed as an approximation (with arbitrary accuracy by choice of m and n) to the SAR test that results when $f(\mathbf{t}, \hat{\gamma}_T)$ is the true joint density of \mathbf{t} under the null hypothesis.

3.2 The semi-parametric case

Even though the DGP is not fully specified in this case, we assume that, under the null hypothesis, the test vector \mathbf{t} has a joint density $f(\mathbf{t}, \gamma)$, which depends on unknown parameter(s) γ . We may not know much about γ including its dimension. We assume there is a bootstrap procedure that allows us to obtain a bootstrap sample of \mathbf{t} values that can be used to approximate the cumulative distribution function (CDF) of the distribution given by $f(\mathbf{t}, \gamma_0)$, where γ_0 is the true unknown value of γ . These bootstrap samples are used to obtain $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_n$ and $\mathbf{t}^{(1)}, \mathbf{t}^{(2)}, \dots, \mathbf{t}^{(n)}$ in either of the test procedures outlined in Section 2.

3.3 Consistency of the estimated p -value

3.3.1 The fully parametric case

Let $\hat{f}_m(\mathbf{t}, \hat{\gamma}_T)$ denote the kernel density estimator of $f(\mathbf{t}, \gamma)$ obtained during the first-round simulation involving m replications, where the bandwidth symbol H is omitted for simplicity. Let p_0 denote the p -value defined through the true density of \mathbf{t} as follows.

$$p_0 = \Pr\{\mathbf{t} : f(\mathbf{t}, \gamma_0) < f(\hat{\mathbf{t}}, \gamma_0) | f(\mathbf{t}, \gamma_0)\}, \quad (9)$$

⁸The requirement that γ be estimated assuming the null hypothesis is true can be relaxed, provided the chosen estimate is consistent under H_0 .

which in our testing procedure, is estimated by $\hat{p}_{T,m}$, the p -value defined via the kernel density estimator under the probability measure implied by the true density of \mathbf{t} . This means that

$$\hat{p}_{T,m} = \Pr \{ \mathbf{t} : \hat{f}_m(\mathbf{t}, \hat{\gamma}_T) < \hat{f}_m(\hat{\mathbf{t}}, \hat{\gamma}_T) | f(\mathbf{t}, \hat{\gamma}_T) \}. \quad (10)$$

In our testing procedure, $\hat{p}_{T,m}$ is approximated by $\hat{p}_{T,m,n}$, the relative frequency of observing $\hat{f}_m(\mathbf{t}^{(i)}, \hat{\gamma}_T) < \hat{f}_m(\hat{\mathbf{t}}, \hat{\gamma}_T)$, for $i = 1, 2, \dots, n$, during the second-round simulation of Step (v).

Assumption 1: $\hat{\gamma}_T$ is a consistent estimate of γ_0 under H_0 .

Assumption 2: $f(\mathbf{t}, \gamma)$ is continuous in γ .

Assumption 3: $f(\mathbf{t}, \gamma)$ as a density function of \mathbf{t} , meets the smoothness conditions given in [Masry \(1996\)](#)⁹.

Assumption 4: The kernel function $K(\cdot)$ and the bandwidth matrix H are chosen to ensure the kernel density estimator is uniformly consistent on a bounded set in which $f(\mathbf{t}, \gamma) > 0$.

Theorem 2: Under Assumptions 1 to 4 in the fully parametric case, as $T \rightarrow \infty$, $m \rightarrow \infty$ and $n \rightarrow \infty$, the estimated p -value denoted as $\hat{p}_{T,m,n}$ is a consistent estimate of the true p -value of our testing procedure.

Proof: The proof of Theorem 2 is given in Appendix 3.

Corollary 2: Under Assumptions 1 to 4 in the fully parametric case and as $T \rightarrow \infty$, $m \rightarrow \infty$ and $n \rightarrow \infty$, our test converges to the SAR test.

Proof: From Theorem 2, as $T \rightarrow \infty$, $m \rightarrow \infty$ and $n \rightarrow \infty$, we have $\hat{p}_{T,m,n} \rightarrow p_0$, and because p_0 is the p -value for the SAR test, the result follows.

Assumptions 1–4 are different to those imposed by Dufour (2006, Section 6). In our set up, his conditions relate to $\hat{f}_{m,H}(\mathbf{t}^{(i)}, \gamma^*)$ and $\hat{f}_{m,H}(\hat{\mathbf{t}}, \gamma^*)$. Our assumption 1 is satisfied by almost all standard econometric estimators, although each comes with its own regularity conditions. Assumptions 2 and 3 impose some restrictions on the type of testing problem we can consider. A problem with verifying these two assumptions is that they involve $f(\mathbf{t}, \gamma)$ which is unknown.

⁹See Theorem 1.4 and its proof of [Li and Racine \(2007\)](#). These conditions are (i) $\inf_{\mathbf{t} \in S} f(\mathbf{t}, \cdot) \geq \delta > 0$, where S is a bounded set excluding the boundary of the support of \mathbf{t} ; (ii) the kernel function $K(\cdot) \in L_1$ satisfies $\|\mathbf{t}\|^2 K(\mathbf{t}) \in L_1$; and (iii) f is bounded and uniformly continuous on R^d , or (iii') the density f is Lipschitz of order τ : $|f(\mathbf{t}, \cdot) - f(\mathbf{u}, \cdot)| \leq C_1 \|\mathbf{t} - \mathbf{u}\|^\tau$, for some $0 < \tau \leq 1$.

Tests of the form of (1) are based on \mathbf{t} having an asymptotic $N(0, V)$ distribution under regularity conditions. These assumptions hold for the $N(0, V)$ density assuming V is finite¹⁰ and a continuous function of γ . So it is likely Assumptions 2 and 3 hold, if T is sufficiently large, for this major class of testing problems. This is all that is needed given that Theorem 2 is an asymptotic result. Our choices of kernel function and bandwidth selectors satisfy Assumption 4.

We have therefore shown that the DSM based on using a consistent estimate of the nuisance parameter γ in order to simulate \mathbf{t} when required, results in an asymptotically justified test. It is also an asymptotic SAR (ASAR) test, where asymptotic means as $T \rightarrow \infty$, $m \rightarrow \infty$ and $n \rightarrow \infty$.

3.3.2 The semi-parametric case

Following Horowitz (2001), let $F_0(\mathbf{y})$ be the CDF of \mathbf{y} under the null hypothesis, and let $G_T(\mathbf{t}, F)$ be the CDF of \mathbf{t} when the sample size is T and F is the exact CDF of \mathbf{y} . Bootstrap sampling involves taking a random sample from some form of empirical distribution for \mathbf{y} . When used for hypothesis testing purposes, it does require the null hypothesis to be imposed in some way. As MacKinnon (2009, p.210) concludes, how well a bootstrap test works depends on how well the bootstrap sampling process “mimics the essential features of the true DGP under the null hypothesis. In general, it seems that bootstrap procedures that are at least partly parametric perform better than ones that are fully nonparametric”. Denote the empirical CDF that the bootstrap sample is taken from as $F_T(\mathbf{y})$. Typically $F_T(\mathbf{y}) \neq F_0(\mathbf{y})$. Hence a bootstrap sample of \mathbf{t} values can be regarded as a random sample from the distribution with CDF $G_T(\mathbf{t}, F_T)$. The bootstrap estimator is consistent if for each $\varepsilon > 0$ and F_0

$$\lim_{T \rightarrow \infty} P_T \left[\sup_{\mathbf{t}} |G_T(\mathbf{t}, F_T) - G_\infty(\mathbf{t}, F_0)| > \varepsilon \right] = 0,$$

where P_T denotes the joint probability distribution of the underlying sample, from which \mathbf{t} is calculated.

Let $\hat{f}_{T,m}^B(\mathbf{t})$ denote the kernel estimator of $f(\mathbf{t}, \gamma_0)$ in the DSM, obtained from the first-round

¹⁰If V is not finite then \mathbf{t} might need to be adjusted by some appropriate function of T to make it finite.

bootstrap simulation involving m replications. Define

$$\hat{p}_{B,T,m} = Pr \{ \mathbf{t} : \hat{f}_{T,m}^B(\mathbf{t}) < \hat{f}_{T,m}^B(\hat{\mathbf{t}}) | G_T(\mathbf{t}, F_T) \}.$$

In the DSM testing procedure, $\hat{p}_{B,T,m}$ is approximated by $\hat{p}_{B,T,m,n}$, the relative frequency of observing

$$\hat{f}_{T,m}^B(\mathbf{t}^{(i)}) < \hat{f}_{T,m}^B(\hat{\mathbf{t}}), \quad \text{for } i = 1, \dots, n,$$

during the second round simulation of Step (v). The test rejects the null hypothesis if $\hat{p}_{B,T,m,n} < \alpha$, where α is the desired significance level.

In order to establish that $\hat{p}_{B,T,m,n}$ is a consistent estimate of p_0 , we need to replace Assumption 1 by the following assumption.

Assumption 1*: The bootstrap samples used to compute $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_m$ and $\mathbf{t}^{(1)}, \mathbf{t}^{(2)}, \dots, \mathbf{t}^{(n)}$ in the DSM are generated by bootstrap sampling that is consistent.

This is not a particularly restrictive assumption although it does require that care is exercised in the selection of the bootstrap procedure used. See Horowitz (2001) for a good discussion of consistency. Horowitz (2018, p.7) observes that “maximum likelihood estimators, generalized method of moments estimators and test statistics are asymptotically linear and asymptotically normal (or are asymptotically Chi-square quadratic forms of asymptotic normal statistics) under the usual regularity conditions. Therefore, the nonparametric bootstrap estimates their asymptotic distributions consistently and can be applied to them”.

Theorem 3: Under Assumptions 1*, 2, 3 and 4 in the semi-parametric case, as $T \rightarrow \infty$, $m \rightarrow \infty$ and $n \rightarrow \infty$, the estimated p -value $\hat{p}_{B,T,m,n}$ is a consistent estimate of the true p -value, p_0 , given by (9).

Proof: The proof of Theorem 3 is given in Appendix 4.

Corollary 3: Under the assumptions of Theorem 3, as $T \rightarrow \infty$, $m \rightarrow \infty$ and $n \rightarrow \infty$, the test based on $\hat{p}_{B,T,m,n}$ converges to the SAR test.

Proof: From Theorem 3, as $T \rightarrow \infty$, $m \rightarrow \infty$ and $n \rightarrow \infty$, we have $\hat{p}_{B,T,m,n} \rightarrow p_0$, and because p_0 is the p -value for the SAR test, the result follows.

Provided t is asymptotically normally distributed, which is typically the case for tests of the form of (1), then given our test converges to the SAR test as $T \rightarrow \infty$, $m \rightarrow \infty$ and $n \rightarrow \infty$, it is also asymptotically equivalent to the test based on (1) from Corollary 1 and is therefore asymptotically pivotal. Because our test uses a bootstrap sample to calculate its p -value (or equivalently its critical value) then it also shares any asymptotic refinements that would come from using bootstrap sampling to find the critical value of an asymptotic test of the form of (1) (see for example, [Horowitz, 2001](#)).

4 Monte Carlo experiments for invariant test statistics

We conducted two separate Monte Carlo experiments in order to study the small sample size and power performance of the new test procedure, where the test vector's null distribution does not depend on nuisance parameters. The testing problems involved are (i) testing for serial correlation in a linear regression model; and (ii) testing for normality in a simple random sample. In each case we compared the performance of different versions of our test procedure, mostly the SSM using the MCMC bandwidth selector and the NRR bandwidth selector with three existing tests. For linear models with just an intercept, we also simulated the DSM versions of our test. In addition, in the case of testing for serial correlation, we explored a bootstrap sampling version of our procedure and we also investigated whether there are issues when d is relatively large.

As simulations of simulations can be very time consuming, we used the following approach to estimate the size and power of the DSM test procedure for the fully parametric case.

- (a) Set parameter values in the DGP under the null hypothesis equal to 1 and repeatedly simulate the DGP to obtain two independent simple random samples of m values of t , denoted as t_1, t_2, \dots, t_m , and $t_1^*, t_2^*, \dots, t_m^*$.
- (b) For the particular choice of bandwidth matrix (NRR or MCMC), compute the value of kernel density given by (3), for t_i^* , namely $\hat{f}_{m,H}(t_i^*)$, for $i = 1, 2, \dots, m$.
- (c) Simulate the DGP under which size or power is to be estimated and calculate a third simple

random sample of n values of t denoted as $t^{(1)}, t^{(2)}, \dots, t^{(n)}$. For each $t^{(j)}$, count the number of $\hat{f}_{m,H}(t_i^*)$ values, for which $\hat{f}_{m,H}(t_i^*) < \hat{f}_{m,H}(t^{(j)})$. The estimated probability of rejection of the null hypothesis is the relative frequency that

$$(\text{count} + 1)/(n + 1) < \alpha,$$

holds for $j = 1, 2, \dots, n$.

For the SSM, only one simple random sample at Step (a) is calculated, denoted as t_1, t_2, \dots, t_m . In place of $\hat{f}_{m,H}(t_i^*)$ in Steps (b) and (c), values of the leave-one-out kernel density $\hat{g}_{m,H}(t_i)$, for $i = 1, 2, \dots, m$, are used.

4.1 Testing for serial correlation of unknown order and form

4.1.1 Experiment design

The first Monte Carlo experiment involves testing the null hypothesis that the disturbances of the classical linear regression model are white noise against the alternative that they contain serial correlation of unknown order and form (see [King, 1987](#)). The null hypothesis is of the form

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (11)$$

where X is $T \times k$, $\boldsymbol{\beta}$ is $k \times 1$ and $\boldsymbol{\varepsilon}$ is $T \times 1$ such that $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I)$. The regressor matrix X is assumed to be fixed or independent of the disturbance vector $\boldsymbol{\varepsilon}$. The alternative is that there is some serial correlation in the components of $\boldsymbol{\varepsilon}$, and it is assumed that it can best be detected by examining r_j , the j th order autocorrelation coefficient, for $j = 1, 2, \dots, d$, where

$$r_j = \frac{\sum_{t=j+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-j}}{\sum_{t=1}^T \hat{\varepsilon}_t^2},$$

in which $\hat{\varepsilon}_t$, $t = 1, 2, \dots, T$, are OLS residuals from (11). In other words,

$$t_j = r_j, \quad \text{for } j = 1, 2, \dots, d,$$

in this problem. Note that r_j is invariant to the values of $\boldsymbol{\beta}$ and σ^2 under (11) and so $f(t)$ does not depend on nuisance parameters.

A standard testing procedure for this problem is the Portmanteau test proposed by [Box and Pierce \(1970\)](#) and extended by [Ljung and Box \(1978\)](#). It involves rejecting the null hypothesis for large values of

$$Q_d = T(T+2) \sum_{j=1}^d \frac{r_j^2}{T-j}. \quad (12)$$

The Monte Carlo experiment involved comparing sizes and powers of the Portmanteau test based on (12) applied using simulated critical values with different forms of the new procedure. Also included in the comparison are two-sided versions of the minimum p -value and product of p -value tests from [Dufour, Khalaf and Voia \(2015\)](#). Because both the Portmanteau test and our new procedure are two-sided in nature, we applied two-sided versions of the minimum p -value and product of p -value tests as follows:

- (i) Assume under H_0 , $\sqrt{T}r_j \sim N(0, 1)$ asymptotically (equation (39) in [Dufour, Khalaf and Voia, 2015](#)). Let $s_j = \sqrt{T}r_j$. If $r_j > 0$, then calculate

$$\delta_j = \Pr(z > s_j | z \sim N(0, 1)),$$

otherwise ($r_j \leq 0$), calculate

$$\delta_j = \Pr(z < s_j | z \sim N(0, 1)),$$

for $j = 1, \dots, d$.

- (ii) The required individual p -values are $p_j = 2\delta_j$, and the test statistic for the minimum p -value test is

$$p = \min_{j=1, \dots, d} p_j,$$

while that of the product of p -values test is

$$q = \prod_{j=1}^d p_j.$$

- (iii) The tests are applied by first calculating p and q (denoted as \hat{p} and \hat{q}) using the actual data and then simulating the calculation of p and q n times under the null hypothesis. The

number of simulated values of p (and q) smaller than \hat{p} (and \hat{q}) is counted (denoted as count), allowing the empirical p -value to be calculated as

$$\hat{p}_n = (\text{count} + 1)/(n + 1).$$

Provided n is chosen so that $\alpha(n + 1)$ is an integer, then these two tests are exact tests at the α level of significance. We set $n = 19,999$.

Let $\mathbf{a}_1 = (1/T)^{1/2}(1, 1, \dots, 1)'$ and

$$\mathbf{a}_i = (2/T)^{1/2}(\cos[\pi(i-1)/(2T)], \cos[3\pi(i-1)/(2T)], \dots, \cos[(2T-1)(i-1)\pi/(2T)])',$$

for $i = 2, 3, \dots, T$. The regressor matrices used are

$X1$ ($T \times 1$): the constant dummy regressor.

$X2$ ($T \times 5$): $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ and \mathbf{a}_5 as regressors.

$X3$ ($T \times 5$): $\mathbf{a}_1, (\mathbf{a}_2 + \mathbf{a}_T)/\sqrt{2}, \dots, (\mathbf{a}_5 + \mathbf{a}_{T-3})/\sqrt{2}$ as regressors.

The choice of $X2$ and $X3$ follows [King \(1985\)](#). The test based on r_1 (see [King, 1981](#)) is approximately UMPI against first-order autoregressive (AR(1)) disturbances for $X2$. $X3$ is known as [Watson's \(1955\)](#) X matrix. Within the class of orthogonal X matrices, this X matrix causes OLS to have minimum efficiency relative to the best linear unbiased estimator for approximately AR(1) disturbances. Thus $X2$ is favourable and $X3$ is unfavourable when it comes to testing for AR(1) disturbances.

Four different forms of the new procedure were used in the simulation for $X1$ with $d = 4$ and $d = 6$, the first two being the SSM and DSM using NRR bandwidth parameters and the second two being SSM and DSM using MCMC bandwidth parameters. In the case of $X2$ and $X3$, only the SSM versions of the new test were used with $d = 4$.

Sizes were calculated by simulating the DGP using (11) with $\varepsilon \sim N(\mathbf{0}, I)$. Powers were calculated for four different DGPs for ε_t in (11), these being ε_t generated by

(i) the stationary AR(1) process given by

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t,$$

or equivalently $(1 - \rho L)\varepsilon_t = u_t$, where $\rho = 0.25$, $u_t \sim IN(0, 1)$ and L is the lag operator;

(ii) the stationary second-order autoregressive (AR(2)) process given by

$$(1 - \rho_1 L)(1 - \rho_2 L)\varepsilon_t = u_t$$

with $(\rho_1, \rho_2) = (0.05, 0.1)$ and $(0.05, 0.2)$ and $u_t \sim IN(0, 1)$;

(iii) the stationary third-order autoregressive (AR(3)) process given by

$$(1 - \rho_1 L)(1 - \rho_2 L)(1 - \rho_3 L)\varepsilon_t = u_t$$

with $(\rho_1, \rho_2, \rho_3) = (0.05, 0.1, 0.15)$ and $(0.05, 0.1, 0.2)$ and $u_t \sim IN(0, 1)$; and

(iv) the stationary fourth-order autoregressive (AR(4)) process given by

$$(1 - \rho_1 L)(1 - \rho_2 L)(1 - \rho_3 L)(1 - \rho_4 L)\varepsilon_t = u_t$$

with $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.05, 0.1, 0.15, 0.15)$, $(0.05, 0.1, 0.05, 0.05)$ and $(0.05, 0.05, 0.05, 0.05)$ and $u_t \sim IN(0, 1)$.

All DGPs were run for $T = 50, 100, 200$ and 500 ; all tests were applied at the 10%, 5% and 1% significance levels and, for the new test procedure, m and n were set to 19,999.

The testing problem in the semi-parametric case involves the null hypothesis of (11) with the only disturbance vector assumptions under the null hypothesis being that $E(\varepsilon) = 0$ and $\text{Var}(\varepsilon) = \sigma^2 I$. Bootstrap sampling for all the tests outlined above was conducted as follows.

(i) First, construct the $T \times (T - k)$ matrix P such that $PP' = M$ and $P'P = I_{T-k}$, where $M = I - X'(X'X)^{-1}X$. The columns of P can be made up of $T - k$ orthonormal eigenvectors corresponding to the $T - k$ unity eigenvalues of the matrix M .

- (ii) Calculate the $(T - k) \times 1$ vector $P'\mathbf{y}$, where \mathbf{y} is the vector of observed responses. Because $P'X = \mathbf{0}$, $P'\mathbf{y} = P'\boldsymbol{\varepsilon}$, so $E(P'\mathbf{y}) = \mathbf{0}$ and $\text{Var}(P'\mathbf{y}) = \sigma^2 I_{T-k}$ under the null hypothesis.
- (iii) Sample $T - k$ times with replacement from the elements of $P'\mathbf{y}$ to obtain a bootstrap sample denoted as $(P'\mathbf{y})^*$, and then form a $T \times 1$ residual vector $\hat{\boldsymbol{\varepsilon}}^* = P(P'\mathbf{y})^*$ by pre-multiplication by P . This residual vector is then used to calculate the required test statistic.

The simulations were run for two sets of errors, the first being normal errors as outlined above and the second being a mixture of two normals given by $0.4N(-2/5, 1) + 0.6N(4/15, 7/3)$, which has zero mean and unity variance, for ε_t under the null and u_t in the alternative processes. This mixture of normals density is bimodal, asymmetric and has heavier tails than the standard normal density.

It is worth noting that when $\varepsilon \sim N(\mathbf{0}, \sigma^2 I)$, the DSM bootstrap version can be regarded as a Monte Carlo test because it fulfills Dufour's (2006) requirements. This follows from viewing $\hat{f}_{T,m}^B(\mathbf{t})$ as our test statistic and observing that

$$\hat{f}_{T,m}^B(\mathbf{t}^{(i)}), \text{ for } i = 1, 2, \dots, n, \text{ and } \hat{f}_{T,m}^B(\hat{\mathbf{t}})$$

all share the same distribution and are exchangeable by construction. The latter result is a consequence of the actual $P'\mathbf{y}$ (and hence $\hat{\mathbf{t}}$) being one of the possible bootstrap samples. Because neither $\hat{f}_{T,m}^B(\cdot)$ nor $\mathbf{t}^{(i)}$, for $i = 1, 2, \dots, n$, depend on unknown parameter values, this shared distribution is independent of unknown parameters. With bootstrap sampling, there is a non-zero probability that two $\hat{f}_{T,m}^B(\mathbf{t}^{(i)})$ values, for $i = 1, 2, \dots, n$, are equal, in which case the procedure would need to be amended as outlined in Dufour's (2006) Proposition 2.3 to make it an exact test provided $\alpha(n + 1)$ is an integer, where α is the desired level of significance. However, as T becomes larger, the number of possible bootstrap samples grows at an extremely fast rate and this probability diminishes to zero very quickly. These observations can also be shown to apply to the minimum p -value and product of p -values tests. They also apply for any value of d , no matter how large and apply to any invariant test of the linear regression based on OLS residuals that shares this null hypothesis and uses this particular form of bootstrap sampling.

4.1.2 Simulation results: Fully parametric case

For the fully parametric case, selected size and power results are presented in Tables 1–3. The following analysis is based on the full set of results that can be found in the online supplementary document. All tests typically have appropriate sizes with the following minor exceptions. For $X1$ when $T = 500$ and $X3$ when $T = 50$, all tests are slightly undersize. While there are minor differences in size accuracy between the SSM and DSM, there is no statistically significant overall difference between the two. In fact, there is no discernible differences in size accuracy between any pair of tests. All tests appear to be doing an excellent job of controlling size.

With respect to power, the most noticeable feature is how much more powerful either version of the new test is than the existing tests when there are five regressors ($X2$ and $X3$). The only exceptions occur when $T = 500$ and for AR(1) disturbances and near AR(1) disturbances. The new test also does very well for $X1$ when $T = 50$ with all versions almost always being more powerful than the best of existing tests. For $X1$ with larger sample sizes, the new test performs relatively better for higher significance levels. There is also a tendency for their relative power performance under $X1$ to be better for $d = 4$ compared to $d = 6$. Across all scenarios, the MCMC based versions are typically slightly more powerful than the NRR based tests although there does not appear to be a systematic difference in power between the SSM and DSM for $X1$.

For regressor matrices $X2$ and $X3$, there is evidence of the existing tests being biased (power lower than size) when $T = 50$ and very occasionally when $T = 100$. The new test does not appear to share this property with its powers always being higher than its sizes. The minimum p -value test does relatively well against the AR(1) disturbances and a near AR(1) process for large T values. Otherwise, it is often the least powerful test with the product of p -values test generally being the best of the existing tests.

Overall, the new procedure has a small power advantage for $X1$, particularly for small sample sizes and larger significance levels. This setting is a case in which standard asymptotic tests perform well. We know that the null distribution of r_j depends heavily on the regressor matrix X . King (1985) provides bounds for the critical values of $2(1 - r_1)$ which he calls the alternative

Durbin-Watson test. The gap between these bounds widens as k increases and it becomes more unlikely that the asymptotic distribution is a useful approximation. The introduction of non-constant regressors (X_2 and X_3) makes the testing problem more complex in terms of its null distribution. The proposed procedure now has a very clear power advantage over the existing tests because of how it is constructed.

4.1.3 Simulation results: Semi-parametric case

For the semi-parametric case, selected size and power results are presented in Tables 4–7. The following analysis is based on the full set of results that can be found in the online supplementary document. There are more occurrences of the estimated size being significantly different from its nominal size, than in the fully parametric case, although the significant differences are relatively small. There are fewer significant deviations for X_1 and X_3 with normal disturbances and more for X_2 and X_3 with non-normal errors. There are no discernable differences in size accuracy between any pair of tests. Overall, while the sizes are not perfect, we judge them to be acceptable.

Turning to power, we see similar patterns to those in the fully parametric case. Both versions of the proposed procedure are typically more powerful than the existing tests for multiple regressors (X_2 and X_3). The only exceptions for X_2 are for AR(1) disturbances and $T = 500$ while for X_3 they extend to $T = 200$ in the AR(1) case as well as a small number of minor occurrences when $T = 500$ for three other disturbance processes. The new test is less dominant for X_1 when $T = 50$ and consequently is the most powerful test a little less often than in the fully parametric case. The test that most often has highest power is still the MCMC version of the new test with its advantage being clearer for non-normal errors. The product of p -values test is second best under this measure.

Again there is evidence of the existing tests being biased when $T = 50$ and occasionally when $T = 100$. This problem appears to be a little more pronounced for X_2 than X_3 . There is no evidence of either version of the new test having this property. Their powers are always higher than their sizes. Overall we can conclude the bootstrap version of the new test has very desirable

power properties although we note its power, relative to that of existing tests, is down slightly compared to the fully parametric case.

4.1.4 Is there a curse of dimensionality?

It is well documented (see [Epanechnikov, 1969](#); [Scott, 2015](#)) that as the dimension, d , of the joint density increases, the difficulty of estimating the density increases. To achieve the same level of accuracy, larger samples may be needed. Fortunately we have control over the number of simulations of \mathbf{t} used to estimate the density and we are only really concerned with one property of the density — its relative height.

In terms of controlling size, large values of d may not be too much of a problem, particularly in the fully parametric case. As we have already observed, when our procedure results in an exact size test, it will be exact no matter how large d is. For the fully parametric case, this includes the DSM version for invariant test statistics and Dufour's MMC version of the DSM test in the non-invariant setting. It also includes our bootstrap version of tests of the standard linear regression without assuming a particular disturbance distribution but when the disturbances just happen to be normally distributed. For non-exact SSM versions of our test, we suspect their sizes might not be too different from those for the equivalent DSM version and therefore not troublesome for large d . Unfortunately, we are not able to make similar conjectures about power.

In order to investigate further whether there is a problem with larger values of d , we ran some extra simulations mostly using the experimental design outlined in Section 4.1.1, but with $T = 100, 200, 300, 500$ and 1000 . The simulations only involved calculating the estimated sizes and powers of NRR-SSM version of the new test, which is non-exact, for $d = 20$ and $d = 50$; in other words for the test based on $r_j, j = 1, 2, \dots, 20$ and also on $r_j, j = 1, 2, \dots, 50$. The DGP was (11) with $X1$ and normally distributed errors. Powers were only calculated for the AR(1) process with $\rho_1 = 0.25$. The number of samples generated in order to estimate $f(\mathbf{t})$ was set in turn to $m = 10,000; 20,000; 30,000; 50,000; 75,000$ and $100,000$. In each case, 20,000 iterations were used to estimate the size or power.

The size results are largely as predicted with estimated sizes being significantly different from their nominal sizes at the 1% level in just 15 out of 150 combinations of T , α and m excluding $m = 10,000$. Of these 15, nine are for $d = 20$ and only six occurred when $d = 50$. So provided 20,000 or more iterations are used, the sizes seem well-behaved. Importantly, there is no evidence of deterioration as d increases.

Selected power results are presented in Table 8. Powers decline as d increases which is to be expected for an alternative of AR(1) errors. Using more estimated correlation coefficients in this case only introduces extra noise into the test vector. Powers change as m changes which in part is caused by sampling error. For $T \geq 200$, these changes are mostly positive as m increases and typically the changes are larger for $d = 50$ than for $d = 20$. For each combination of T and α for which $T \geq 200$, the largest or second largest estimated power occurs when $m = 100,000$ with the only exceptions being for $T = 200$ and $\alpha = 0.05$. This suggests that typically one can get a power improvement by setting $m = 100,000$ rather than 20,000 unless the power is already close to or at its maximum of one. In general, much of this improvement can be achieved with $m = 75,000$.

There may be computational reasons which make it difficult to set $m = 75,000$ or 100,000, in which case, we suggest starting with $m = 20,000$ and only if the estimated p -value is judged to be close to α and computational resources allow, then continue sampling (effectively increasing m) until $m = 40,000$ when a second decision of whether to continue sampling can be made. This iteration process could be continued until $m = 100,000$.

4.2 Testing for normality

4.2.1 Experimental design

In many statistical situations, random observations are often assumed to be normally distributed for the purpose of statistical inference. Therefore, it is important to be able to test for normality (see, for example, [Shapiro and Wilk, 1965](#); [D'Agostino, 1971, 1972](#); [Bowman and Shenton, 1975](#); [Pearson, D'Agostino and Bowman, 1977](#); [Jarque and Bera, 1980, 1987](#); [Spiegelhalter, 1980](#); [Thode, 2002](#); [Dufour, Khalaf and Beaulieu, 2003](#); [Dufour, Farhat and Khalaf, 2004a](#)).

The second Monte Carlo experiment involved the problem of testing the null hypothesis that a simple random sample is independently and identically normally distributed with unknown mean (μ) and unknown variance (σ^2) against the alternative that it is non-normally distributed. In other words, (11) with $X = X1$ is the model under the null hypothesis. Evidence of non-normality is often obtained from sample measures of skewness and kurtosis denoted as $\sqrt{b_1}$ and b_2 , respectively, where

$$b_1 = \hat{\mu}_3^2 / \hat{\mu}_2^3, \quad b_2 = \hat{\mu}_4 / \hat{\mu}_2^2,$$

and $\hat{\mu}_i = \sum_{t=1}^T (y_t - \bar{y})^i / T$, for $i = 2, 3, 4$, with $\bar{y} = \sum_{t=1}^T y_t / T$. Jarque and Bera (1980, 1987), D'Agostino and Stephens (1986), Urzúa (1996) and Thode (2002) have discussed omnibus tests for normality that combine information from $\sqrt{b_1}$ and b_2 , and Monte Carlo tests have been recommended by Dufour, Farhat, Gardiol and Khalaf (1998), Dufour and Khalaf (2001), Dufour, Khalaf and Beaulieu (2010) and Dufour, Farhat and Khalaf (2004a). As Dufour and Khalaf (2001) observe, the joint distribution of $\sqrt{b_1}$ and b_2 under the null hypothesis of normality, is independent of nuisance parameters

This experiment involves comparing the small-sample properties of four versions of our test procedure based on the test vector $(\sqrt{b_1}, b_2)'$ in which $\sqrt{b_1}$ takes the sign of $\hat{\mu}_3$, with the Jarque-Bera LM test (Jarque and Bera, 1980, 1987) whose test statistics is

$$JB = T \left[\frac{(\sqrt{b_1})^2}{6} + \frac{(b_2 - 3)^2}{24} \right]. \quad (13)$$

The versions of the new test procedure used are the four combinations of the DSM and SSM with the NRR or MCMC bandwidth selector. Note that the DSM yields an exact test in this case because n was set to 19,999.

A small number of simulation studies have revealed that the size of the JB test is incorrect for small- and moderate-sized samples particularly in the context of the linear regression model. A more straightforward solution is to use Monte Carlo simulations to obtain correct critical values, which is the approach we used (see, for example, Dufour and Khalaf, 2001; Poitras, 2006).

Also included in the comparison were two-sided versions of the minimum p -value and

product of p -value tests from [Dufour, Farhat and Khalaf \(2004a\)](#) and [Dufour \(2006\)](#). We used

$$\sqrt{T/6}\sqrt{b_1} \sim N(0,1) \text{ and } \sqrt{T/24}(b_2 - 3) \sim N(0,1), \text{ asymptotically,}$$

in order to calculate the individual p -values required for these tests.

Sizes were calculated by simulating (11) with $X = X1$ and $\varepsilon_t \sim IN(0,1)$. Powers were calculated for four alternative distributions for ε_t , these being the Student's t distribution with 5 degrees of freedom denoted t_5 , the χ^2 distribution with 3 degrees of freedom denoted χ_3^2 , the Gamma distribution with shape and scale parameters being 2 and 1, and the log normal distribution with mean zero and standard deviation 0.5. These tests were compared for sample sizes of $T = 30, 50, 75$, and 100 with the values of m and n both being 19,999.

4.2.2 Simulation results

The size and power results are given in Table 9. All seven tests have excellent sizes, and there are no discernible differences between the estimated sizes of the seven tests, including between the SSM and DSM versions of our test.

With regards to power, a version of the new test procedure is more often than not the most powerful (or equal most powerful) of the seven tests. Across the 48 combinations of α , T and non-normality, an existing test is more powerful than every version of the new test on only 14 occasions. Unlike for the previous case of testing for autocorrelation, no particular version of the new test dominates. Collectively, the relative power of the new tests are best against t_5 and χ_3^2 distributions and weakest against the log-normal distribution, particularly for $T = 30$ and 50. Against all distributions, the relative power of the new test is poorest for $T = 30$ and for $\alpha = 0.10$ when the true distribution is non-symmetric. With a few exceptions, the product of p -values test is typically the least powerful of the seven tests for $\alpha = 0.01$ and 0.05, while the minimum p -value test is generally the least powerful for $\alpha = 0.10$ and $T \geq 50$, except against the log-normal distribution. The JB test has best power against non-symmetric distributions for larger T values and $\alpha = 0.10$,

Out of the three existing tests, the minimum p -value test has the highest power almost always for $\alpha = 0.01$ and against non-symmetric distributions for $\alpha = 0.05$. The JB test is typically the most powerful of the three when $\alpha = 0.10$. The product of p -values test also performs relatively well for $\alpha = 0.10$ and has good relative power for $\alpha = 0.05$ against the t_5 distribution.

5 Information matrix test

It is often important to test whether a model is correctly specified. [White \(1982\)](#) showed that when a model is correctly specified and estimated by maximizing the likelihood function, the information matrix should be asymptotically equal to the negative Hessian matrix. The information matrix test introduced by [White \(1982\)](#), aims to test the significance of the discrepancy between the negative Hessian and the outer product of the score vector, where the lower triangular components of the matrix of such differences are organized into a vector which we call the test vector in this paper. [Chesher \(1984\)](#) showed that the IM test can be viewed as an LM test for specification error against the alternative of parameter heterogeneity. [Chesher \(1983\)](#) and [Lancaster \(1984\)](#) presented an TR^2 version of the IM test, where T is the sample size and R^2 is the goodness of fit obtained through the OLS regression of a column of ones on a matrix whose elements are functions of the first and second derivatives of the log-likelihood function. For the normal fixed regressor linear model, [Hall \(1987\)](#) showed that the LM version of the IM test can be asymptotically decomposed into the sum of three components, where one is the general test for heteroscedasticity proposed by [White \(1982\)](#), and the other two components aim to test for normality in different ways.

The use of the IM test in applied econometrics is limited because its actual size based on the asymptotic critical value often differs greatly from its nominal size. This phenomenon has been detected in Monte Carlo experiments conducted by [Taylor \(1987\)](#), [Orme \(1990\)](#), [Chesher and Spady \(1991\)](#) and [Davidson and MacKinnon \(1992\)](#). [Davidson and MacKinnon \(1992\)](#) proposed to deal with this problem by using double-length artificial regressions to compute a variant of the IM test statistic, but models for discrete, censored or truncated data cannot be dealt with

via this method. [Chesher and Spady \(1991\)](#) suggested obtaining the critical value for the IM test from an Edgeworth expansion of the finite-sample distribution of the test statistic. Their Monte Carlo investigation indicates that such a critical value provides a good approximation to the true critical value. In the examples considered by [Chesher and Spady \(1991\)](#), the Edgeworth expansions are independent of the parameters of the models being tested, and therefore, the IM test statistic is pivotal ([Horowitz, 1994](#)). However, this is not a general case, and it is often very difficult to decide whether the IM test statistic is pivotal.

[Horowitz \(1994\)](#) proposed a bootstrapping procedure to obtain critical values for the IM test and showed that one can easily obtain good finite-sample critical values for the IM test through bootstrapping rather than through Edgeworth expansions or other algebraically complicated manipulations. Moreover, he discussed the power performance of three versions of the IM test through Monte Carlo simulation. His results showed that all three have much lower powers for size-corrected critical values than those computed according to asymptotic critical values. Therefore, it seems that getting the size right and achieving higher power are different tasks.

Most existing versions of the IM test rely on an estimate of the asymptotic covariance matrix of the test vector. The analytical form of the asymptotic covariance matrix is complicated and involves the third derivative of the log-likelihood function. [Lancaster \(1984\)](#) showed that the matrix can be estimated without calculating the third derivative. [Dhaene and Hoorelbeke \(2004\)](#) indicated that the incorrect-size problem results from the inaccurate estimate of the covariance matrix. They proposed estimating the covariance matrix of the test vector through parametric bootstrapping.

Let $f(y|\theta)$ denote the density for a postulated model where θ is an $r \times 1$ vector of parameters. Let $\mathbf{y} = (y_1, y_2, \dots, y_T)'$ be the vector of observations, and $\ell(\mathbf{y}|\theta) = \log f(\mathbf{y}|\theta)$ the logarithmic density. We introduce the following notation.

$$A(\theta) = E \left[\frac{\partial^2 \ell(\mathbf{y}|\theta)}{\partial \theta \partial \theta'} \right], \quad A_T(\mathbf{y}, \theta) = \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 \ell(y_t|\theta)}{\partial \theta \partial \theta'},$$

$$B(\theta) = E \left[\frac{\partial \ell(\mathbf{y}|\theta)}{\partial \theta} \frac{\partial \ell(\mathbf{y}|\theta)}{\partial \theta'} \right], \quad B_T(\mathbf{y}, \theta) = \frac{1}{T} \sum_{t=1}^T \frac{\partial \ell(y_t|\theta)}{\partial \theta} \frac{\partial \ell(y_t|\theta)}{\partial \theta'},$$

where expectations are taken with respect to the true density. When the model is correctly specified, the true density is $f(\mathbf{y}|\theta)$. Let θ_0 be the true value of θ .

The IM test is based on the information-matrix equality, which states that $A(\theta_0) + B(\theta_0) = 0$ when the model is correctly specified. Given the vector of T independent observations, \mathbf{y} , the IM test investigates the statistical significance of $A_T(\mathbf{y}, \hat{\theta}) + B_T(\mathbf{y}, \hat{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimator of θ .

Let \mathbf{t} denote the vector of indicators (test vector) whose elements are D_{ij} , for $i = 1, 2, \dots, r$, and $j = 1, \dots, i$, where

$$D_{ij} = \frac{1}{T} \sum_{t=1}^T \left[\frac{\partial \ell(y_t|\theta)}{\partial \theta_i} \frac{\partial \ell(y_t|\theta)}{\partial \theta_j} + \frac{\partial^2 \ell(y_t|\theta)}{\partial \theta_i \partial \theta_j} \right] \Big|_{\theta=\hat{\theta}}. \quad (14)$$

White (1982) shows that under regularity conditions, the IM test statistic is of the form

$$\xi = T \mathbf{t}' \hat{V}^{-1} \mathbf{t}, \quad (15)$$

where \hat{V} is the consistent estimator of the covariance matrix of \mathbf{t} under H_0 . Under the null hypothesis, ξ is distributed asymptotically χ_q^2 with $q = r(r+1)/2$, and this is based on the asymptotic null distribution of $\mathbf{t} \sim N(0, V(\theta))$. The IM test statistic depends on the estimate of the covariance matrix which in turn depends on the estimate of θ .

Our proposed IM testing procedure aims to estimate the joint density of the test vector, \mathbf{t} , and does not depend on \hat{V} . We believe basing the test directly on the vector \mathbf{t} and not needing an estimate of the covariance matrix of \mathbf{t} under H_0 will give our proposed test a clear advantage. Not having to assume the mean of \mathbf{t} is zero is another advantage. However, the null distribution of \mathbf{t} is highly dependent on θ , the unknown model parameters.

Since $A_T(\mathbf{y}, \hat{\theta}) + B_T(\mathbf{y}, \hat{\theta})$ is a symmetric matrix, a test of the complete IM identity can be based on the lower triangular elements of $A_T(\mathbf{y}, \hat{\theta}) + B_T(\mathbf{y}, \hat{\theta})$ (or D_{ij}). However, according to White (1982), in many situations it is inappropriate to base the test on all q indicators because some indicators may be identically zero, or linear combinations of other indicators. In either case, it is appropriate to ignore such indicators. In the remainder of this paper, the IM tests are based on the maximum number of linearly independent non-zero indicators.

6 Simulation study of the new test applied to the IM test

This section reports a Monte Carlo simulation study which aims to compare the finite-sample size and power performance of our proposed SSM based on t with the [Lancaster \(1984\)](#) form of the IM test denoted by IM_L and the [Dhaene and Hoorelbeke \(2004\)](#) form of the IM test denoted by IM_{DH} . We also included two versions of both the minimum p -value (denoted as $\min p_L$ and $\min p_{DH}$) and the product of p -value tests (denoted as prod_L and prod_{DH}) based on the components of t . The test vector t is assumed to follow a multivariate normal distribution with mean zero and covariance matrix being either calculated by the [Lancaster \(1984\)](#) method in the L version and the [Dhaene and Hoorelbeke \(2004\)](#) method in the DH version. This allows the calculation of p -values for two-sided versions of the component tests as outlined in Section 4.1.1.

6.1 Null and alternative hypotheses

The study covers two different settings. The null hypothesis in the first setting is the normal linear regression model given by

$$y_t = x_t' \beta + u_t, \quad (16)$$

for $t = 1, 2, \dots, T$, where $u_t \sim IN(0, \sigma^2)$, x_t is a $k \times 1$ vector of regressors, and β is a $k \times 1$ vector of parameters. Following [Dhaene and Hoorelbeke \(2004\)](#), we examine the power of the IM test under the heteroscedastic alternative of

$$y_t = x_t' \beta + u_t, \quad u_t \sim N(0, |x_t' \beta|), \quad \text{for } t = 1, 2, \dots, T. \quad (17)$$

In the second setting, the null model is the Tobit model given by

$$y_t = \begin{cases} x_t' \beta + u_t & \text{if } x_t' \beta + u_t > 0 \\ 0 & \text{if } x_t' \beta + u_t \leq 0 \end{cases}, \quad (18)$$

for $t = 1, 2, \dots, T$, $u_t \sim IN(0, \sigma^2)$, x_t is a $k \times 1$ vector of regressors, and β is a $k \times 1$ vector of parameters.

Following [Horowitz \(1994\)](#), we examine the power of the IM tests under the models given by

$$y_t = \max(0, x_t' \beta + u_t), \quad u_t \sim IN(0, \exp(0.5 x_t' \beta)), \quad (19)$$

and

$$y_t = \max(0, x_t' \beta + 0.75x_{t,2}x_{t,3} + u_t), \quad u_t \sim IN(0, 1), \quad (20)$$

for $t = 1, 2, \dots, T$, where $x_{t,2}$ and $x_{t,3}$ are the two non-intercept components of x_t . Note that model (19) involves a heteroscedastic alternative while model (20) has an incorrect mean function.

The experiments consist of applying various forms of IM tests along with the proposed method of testing to the linear regression and Tobit models. In both models, x_t consists of an intercept component and either one or two additional variables. The values of x_t are fixed in repeated samples. The values of the β vectors are $(0.75, 1)'$ and $(0.75, 1, 1)'$ when x_t is 2×1 and 3×1 , respectively. The non-intercept components of x_t are sampled independently from the standard normal distribution. The value of σ^2 is set to 1 in all of the experiments. The sample sizes are 50, 100, 200 and 300, and we set $n = m = 10,000$. The t vector is 5×1 in the one-regressor case and 9×1 in the two-regressor case. For the IM_{DH} test statistic and the DH versions of the minimum p -value and product of p -value tests, we used 50 parametric bootstrap samples to estimate the covariance matrix, \hat{V} , following [Dhaene and Hoorelbeke \(2004\)](#). Size-corrected critical values, which were obtained via simulation under the null hypothesis with known true parameters, were used for computing the powers of the IM_L and IM_{DH} tests. Note that when the IM test statistic is not pivotal (i.e. for the Tobit model), these size-corrected critical values for the IM_L and IM_{DH} tests cannot be calculated in a practical application because the true parameter values under the null hypothesis are unknown. We have used these critical values in the simulation so that the powers of the respective tests can be compared fairly.

In a separate simulation using recommended χ^2 critical values, we found estimated sizes for the IM_L test and $\alpha = 0.10, 0.05$ and 0.01 of 0.712, 0.611 and 0.417 respectively, when $T = 50$ and 0.553, 0.462 and 0.298 when $T = 100$. For the IM_{DH} test we did see an improvement in these estimated sizes to, respectively, 0.164, 0.114 and 0.064 when $T = 50$ and for $T = 100$, to 0.151, 0.104 and 0.048. All these estimated sizes are highly significantly larger than their nominal sizes.

Clearly the use of χ^2 critical values for these two tests is not appropriate, particularly if we wish to compare powers.

6.2 Simulation results from the linear model

The results for the linear model are presented in Tables 10 and 11 for sizes and powers, respectively. From Table 10, we see that the sizes for the proposed test and the various versions of the IM test are reasonably close to their corresponding nominal sizes. There are a sprinkling of the estimated sizes that are significantly different from their nominal sizes but the differences are typically small and not particularly concerning. Their occurrence is greater for the two regressor case when $d = 9$. The minp_{DH} test has the best sizes overall with the remaining tests being remarkably similar in their performance.

From Table 11, we see that the proposed method is always more powerful, typically by a large margin, than all the other tests, with the exception of when both the new test and the prod_{L} test have power of one. We see settings in which the new test provides four or five-fold increases in power, although this advantage declines as the powers approach one. There are many examples, where the use of the new test can more than double the power of an existing test. For $T = 100, 200$ and 300, the prod_{L} test is typically the second most powerful test while for $T = 50$, it is the IM_{DH} test. The IM_{DH} test is almost always more powerful than the IM_{L} test which is generally the least powerful of all the tests and shows evidence of being biased when $T = 50$. The dominance of the IM_{DH} test over the IM_{L} test suggests that there is a power advantage in using bootstrap methods to estimate the covariance matrix, although this observation does not extend to the minimum p -value and product of p -value tests.

6.3 Simulation results from the Tobit model

The size and power results for the Tobit model are presented in Tables 12–13. In Table 12, we see that the estimated sizes of the proposed test are not significantly different at the 1% level to their nominal sizes for both the one-regressor and two-regressor models although there are three

occurrences of a 5% significant difference in the latter case. The sizes of the minp_{DH} and prod_{DH} tests show similar behaviour. On the other hand, the estimated sizes for the IM_{L} , IM_{DH} and minp_{L} tests show a greater number of significant deviations from their nominal sizes although none are particularly troublesome.

Table 13 presents the estimated powers of the tests, when model (19) is the true alternative hypothesis. The new test is always more powerful than the other six existing tests. The biggest improvements in power occur for smaller sample sizes and smaller significance levels when the new test often has more than double the power of the best existing test. The next best test is almost always a DH based test, mostly likely the IM_{DH} test although the minp_{DH} test performs well in the one regressor case.

The last column panel of Table 13 presents the estimated powers of the tests when model (20) is the true model. With the exception of a couple of cases when $\alpha = 0.01$ and the sample size is small, the new test is always the most powerful of the seven tests with the IM_{L} or IM_{DH} test typically being the second most powerful.

Overall, the results of our simulation study suggest that at least in the context of IM testing, the new test has excellent finite sample size and power properties. There is almost no evidence to suggest the sizes are different from their nominal sizes and in some cases the powers can be more than double those of the existing tests. The requirement in the two-regressor model case of needing to estimate a nine-dimensional joint density function does not seem to hinder the performance of the proposed test.

7 Test for misspecification of cross-market prediction models

Stock market analysts have found that the Australian stock market often follows the overnight US stock market, and sometimes major European stock markets as well. Therefore, it is possible to forecast the Australian stock market daily return based on the overnight stock market returns on the US stock index and a major European stock index. Such a cross-market predictive relationship is one type of cross-market relationship, which is a broad area that has been extensively studied

in the finance literature (see, for example, [Forbes and Rigobon, 2002](#); [Longstaff, 2010](#); [Rapach, Strauss and Zhou, 2013](#); [Bekaert, Ehrmann, Fratzscher and Mehl, 2014](#)). In order to model such a predictive relationship, one possible model is

$$y_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 z_{t-1} + \varepsilon_t, \quad (21)$$

for $t = 1, 2, \dots, T$, where y_t is the daily return on the Australian S&P/ASX 200 index, and x_{t-1} and z_{t-1} are the lagged US S&P 500 and German DAX daily returns standardized by the corresponding daily VIX measures, respectively.¹¹

In order to capture the stylized facts of y_t , one reasonable specification is to assume that ε_t , for $t = 1, 2, \dots, T$, are conditional heteroscedastic. Another is to standardize y_t by its daily VIX and assume that ε_t , for $t = 1, 2, \dots, T$, are iid $N(0, \sigma^2)$. The latter specification might be inappropriate when there exists a remaining GARCH effect even after the standardization of y_t . Nonetheless, the latter specification can be tested through each of the three versions of the IM test discussed in the previous section. This might be considered a particular appropriate test for misspecification given that [Hall \(1987\)](#) showed that the LM version of the IM test in this setting can be asymptotically decomposed into the sum of three components, one of which tests for heteroscedasticity and the other two test for non-normality.

We collected a sample of these three stock daily returns from the 5th January 2010 to the 31st July 2014 with the sample size being 1110. All three daily return series were standardized by their respective daily VIX measures and the errors, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T$, were assumed to be iid $N(0, \sigma^2)$. The model was estimated using OLS, and the parameter vector $(\beta_0, \beta_1, \beta_2, \sigma)$ estimate was (0.0054, 0.7879, 0.1295, 0.6331). Applying the LM test for GARCH effects to the residuals¹², we find that the p -values are 0.0000 for lags up to five, 0.0002 for lags up to ten, and 0.0058 for lags up to twenty. Thus, the LM test rejects the null hypothesis of no GARCH effect in the residual series of (21) at the 1% significance level.

¹¹If the regressors are not standardized, the cross-market linear relationship is very likely to be twisted by high volatilities, which usually occur when stock returns are extremely negative or positive values.

¹²See, for example, [Engle \(1982\)](#) and [Lee and King \(1993\)](#) for discussion of the Lagrange multiplier test.

According to this finding, the model given by (21) with iid $N(0, \sigma^2)$ errors is likely to be misspecified. The IM test based on our testing procedure, as well as the other two versions of the IM test, were carried out to test for such misspecification. With respect to Assumptions 1–4 needed for the p -value of our procedure to converge to that of the exact SAR test, we note that under the null hypothesis and conditional on the values of the regressors, OLS provides consistent estimates and the null model satisfies the regularity conditions of White’s (1982) IM test.¹³ Consequently Assumptions 2 and 3 hold for sufficiently large T which, as noted in Section 3.3.1, is all that is needed. Our choice of bandwidth matrix satisfies Assumption 4.

Our testing procedure produces a p -value of 0.0094, while the p -values of IM_{DH} and IM_L tests are 0.0073 and 0.3573, respectively. Thus, our test and the IM_{DH} test both reject this specification at the 1% significance level. However, the Lancaster version of the IM test does not reject it.

Investors sometimes may be interested in how positive daily returns in the Australian stock market are affected by the overnight daily returns in the US market and a major European market. Such a cross-market relationship might be explained by a Tobit model:

$$y_t = \begin{cases} \beta_0 + \beta_1 x_{t-1} + \beta_2 z_{t-1} + u_t & \text{if } \beta_0 + \beta_1 x_{t-1} + \beta_2 z_{t-1} + u_t > 0 \\ 0 & \text{if } \beta_0 + \beta_1 x_{t-1} + \beta_2 z_{t-1} + u_t \leq 0 \end{cases}, \quad (22)$$

where u_t , for $t = 1, 2, \dots, T$, are iid $N(0, 1/h^2)$. Note that there is no evidence supporting the Tobit specification of this model. Using the maximum likelihood estimation method, we fitted the model to a shorter sample of daily returns from the 3rd January 2013 to the 31st July 2014, a sample of 382 observations. The parameter vector $(\beta_0, \beta_1, \beta_2, h)$ was estimated as (0.0075, 0.8752, 0.0879, 1.4937).

To test whether the Tobit specification is misspecified, we carried out the IM test based on our testing procedure together with IM_{DH} and IM_L tests. Our estimator, under the null hypothesis and conditional on the regressors, is consistent, which means Assumption 1 holds. Assumptions 2 and 3 will hold at least asymptotically under the same regularity conditions as required for White’s (1982) IM test to be asymptotically valid. Our choice of bandwidth

¹³See Hall (1987) for a comprehensive discussion of the IM test in the context of the standard linear regression model.

matrix satisfies Assumption 4. The p -value of our test is 0.0001, and the null hypothesis of no misspecification is rejected. However, the p -values of the IM_{DH} and IM_L tests are respectively, 0.1312 and 0.3989, which do not reject the null hypothesis at significance levels of 10% and less.

8 Conclusion

This paper presents a new procedure for hypothesis testing based on a vector of statistics. It involves simulating the statistics under the null hypothesis and then estimating their joint density using a multivariate kernel density estimator. This allows the p -value of the vector of statistics to be estimated giving rise to an approximate SAR test. Simulation is straight forward when the null DGP is fully parameterized and the null distribution of the vector of statistics is independent of nuisance parameters. It also allows us to approximate the SAR test with arbitrary accuracy by our choice of the number of iterations in the required simulations. If the model under the null hypothesis is semi-parametric, then a bootstrap option for simulating the DGP under null is available. Furthermore, it is possible to conduct the test in such a way that it is an exact size test. In the fully parametric case where the null distribution of the test vector is dependent on nuisance parameters, these parameters are first estimated and then used to simulate the vector of statistics under the null hypothesis. Our test can be viewed as an approximation (with arbitrary accuracy) to the SAR test that results when the estimated nuisance parameters are equal to the true nuisance parameters under the null hypothesis. While not totally ideal, this is an approximation to an appropriate finite sample test rather than an asymptotic test which is the focus of much of the literature on tests of the form of (1). Bootstrapping can be used for semi-parametric models. We prove that the resultant p -value in both the fully parametric and semi-parametric cases is a consistent estimate of the p -value of the exact SAR test under some regularity conditions and assuming consistency of the nuisance parameter estimates or the bootstrap sampling procedure.

The small-sample properties of the proposed procedure were investigated via simulation in the context of testing for autocorrelation, testing for normality, and testing for model misspecifi-

cation through the information matrix. We find that our testing procedure has appropriate sizes and that powers that are typically better than, or in some cases as good as those of existing tests. It appears that for relatively simple testing problems with few nuisance parameters, such as testing for autocorrelation and non-normality in a random sample, the new procedure typically has a slight advantage in terms of power. We see clear evidence of that advantage increasing as we turn to the more difficult problem of testing for serial correlation in the linear regression model and testing for misspecification via the information matrix in the linear regression model and the Tobit model.

The standard approach in these more difficult testing problems, is to derive the asymptotic distribution of the vector of statistics, estimate the asymptotic covariance matrix and calculate the usual quadratic form that has an asymptotic χ^2 distribution under the null hypothesis and assuming regularity conditions hold. Each of the steps of test construction involves an approximation that has the potential to affect the power of the resultant test. This includes approximating the mean of t , its covariance matrix and its joint distribution. Our approach focuses directly on estimating the small-sample null distribution of the vector of statistics in order to estimate the overall p -value. When that distribution is known, our test has the smallest acceptance region of tests of that size in the d -dimensional sample space of t .

The problem of testing for serial correlation in the linear regression model is one in which the mean of the test vector under the null can be rather different from its asymptotic mean. This has been identified as a possible cause of local biasedness in the Wald test. The use of an approximate SAR test can help mitigate against this problem. The simulation results for the linear regression model with five regressors show local biasedness in the power of existing tests, typically for $T = 50$ with no evidence of biasedness in the new test. This is true for both the fully parametric case (Gaussian errors) and the semi-parametric case (unknown error distribution). It occurred despite all the existing tests having exact or approximately exact critical values. This illustrates a potential problem for asymptotic tests of the form (1) that deserves more attention.

We simulated a bootstrap version of the procedure in the case of testing for autocorrelation

in the linear regression. For this setting, we did find that the DSM version of our test can be exact when the disturbances just happen to be normally distributed.¹⁴ Presumably, different error distributions may result in different sizes but we can take some comfort in the fact that sizes can be exact for one important class of distributions. The simulation results in the constant mean case do suggest a slight deterioration in the relative power of the new test for the smallest sample size of $T = 50$. It does seem that a reasonable sample size is needed to get the best from using the bootstrap option.

There is a lot of discussion in the bootstrap literature about the need for statistics to be pivotal or asymptotically pivotal. This allows the bootstrapped test to benefit from a higher-order improvement in the accuracy of its size. A statistic of the form of (1) is asymptotically pivotal while t typically may not be. We have been able to infer that under regularity conditions, our test statistic based on the estimated joint density of t is asymptotically pivotal. It therefore shares any asymptotic refinements that would come from using bootstrap sampling to find the critical value of a test of the form of (1).

An important step in our procedure is the selection of bandwidth values for kernel density estimation. We found that the MCMC based approach is slightly better than the NRR particularly for large values of d , although there is a very big difference in the computational time required. So if this is an issue then the use of the NRR can provide very acceptable results.

There is a question of how well the procedure might work when the dimension of the test vector, d , is large. Whenever our procedure provides an exact test, the test will be exact no matter how large d is. This observation gives us some confidence that our procedure will have reasonable appropriate sizes for large d values, more so for the fully parametric case than the semi-parametric case. With respect to power, we have found some evidence of power increasing mildly as the number of simulations used to estimate the joint density increases, with these power improvements being more pronounced for large d values.

¹⁴We could have made the more general assumption that the vector ε follows a spherically symmetric distribution with a joint density of the form $\phi(\varepsilon'\varepsilon)$ which includes $N(0, \sigma^2 I_n)$ as a special case. As King (1981) notes, all statistics that are invariant to the scale of ε have the same distribution for all spherically symmetric distributions including normality.

We applied three versions of the information test for misspecification to some cross-market prediction models of daily returns on the Australian stock market. We find the three tests can have quite different estimated p -values which can result in different test outcomes. The results of tests for GARCH effects in the model's residuals tend to support the reliability of our proposed test procedure over the other two tests.

Acknowledgements

We are very grateful to the Editor, Tong Li, and two reviewers for some very constructive suggestions which helped improve the paper and our understanding of the approach we are proposing. We also thank Graham Elliott, Jiti Gao, Rob Hyndman, Oliver Linton, Zen Lu, Esfandiar Maa-soumi, Hashem Pesaran, Aman Ullah, Kees Jan van Garderen and Halbert White for their useful comments and suggestions. This research was supported under Australian Research Council's *Discovery Projects* funding scheme. An early version of this paper was presented at the Econometric Society World Congress held in Shanghai in August 2010 and was entitled "A new procedure for multiple testing of econometric models".

Appendix 1: Proof that the proposed test for $\beta_2 = 0$ in (2) based on $\hat{\beta}_2$ is equivalent to the F test

Because under H_0

$$t = \hat{\beta}_2 \sim N\left(0, \sigma^2 (X_2' M_1 X_2)^{-1}\right),$$

then

$$f(t, \sigma^2) = |2\pi\sigma^2 (X_2' M_1 X_2)^{-1}|^{-1/2} \exp[-t'(X_2' M_1 X_2)t/(2\sigma^2)].$$

Our proposed test is based on the estimated p -value

$$\begin{aligned} \hat{p} &= Pr \{t: f(t, \hat{\sigma}^2) < f(\hat{t}, \hat{\sigma}^2)\} \\ &= Pr \{t: \exp[-t'(X_2' M_1 X_2)t/(2\hat{\sigma}^2)] < \exp[-\hat{t}'(X_2' M_1 X_2)\hat{t}/(2\hat{\sigma}^2)]\}. \end{aligned}$$

This p -value is equivalent to the p -value of the test which rejects H_0 for

$$\exp[-\mathbf{t}'(X_2' M_1 X_2) \mathbf{t} / (2\hat{\sigma}^2)] < c,$$

or equivalently rejects H_0 for

$$\mathbf{t}'(X_2' M_1 X_2) \mathbf{t} / (2\hat{\sigma}^2) > c^*,$$

or equivalently rejects H_0 for

$$F = \frac{\mathbf{y}'(M_1 - M) \mathbf{y} / d}{\mathbf{z}' \mathbf{z} / (T - k)} > F_{1-\alpha}(d, T - k),$$

where $M = I - X(X'X)^{-1}X'$, $F_{1-\alpha}(d, T - k)$ is the $(1 - \alpha)100$ percentile of the central F distribution with d and $(T - k)$ degrees of freedom, and α is the desired level of significance. This is the well known F test. The last equivalence follows because through substitution and algebra,

$$\mathbf{t}'(X_2' M_1 X_2)^{-1} \mathbf{t} = \mathbf{y}' M_1 X_2 (X_2' M_1 X_2)^{-1} X_2' M_1 \mathbf{y},$$

and we can show that

$$M_1 X_2 (X_2' M_1 X_2)^{-1} X_2' M_1 = M_1 - M.$$

Appendix 2: Proof of Theorem 1

Proof: Given $\mathbf{t} \sim N(\mathbf{a}, W)$, rejecting the null for small values of

$$f(\hat{\mathbf{t}}) = (2\pi)^{-d/2} |W|^{-1/2} \exp\{-\hat{(\mathbf{t}} - \mathbf{a})' W^{-1} (\hat{\mathbf{t}} - \mathbf{a}) / 2\}$$

is equivalent to rejecting the null for small values of

$$\exp\{-\hat{(\mathbf{t}} - \mathbf{a})' W^{-1} (\hat{\mathbf{t}} - \mathbf{a}) / 2\}$$

or equivalent to rejecting the null for large values of

$$(\hat{\mathbf{t}} - \mathbf{a})' W^{-1} (\hat{\mathbf{t}} - \mathbf{a}),$$

which completes the proof. \square

Appendix 3: Proof of Theorem 2

For multivariate kernel density estimation, [Li and Racine \(2007\)](#) showed that under the smoothness conditions on the true density given in [Masry \(1996\)](#), the kernel density estimator is uniformly consistent on a bounded set, in which the true density is greater than zero. This implies that for any γ value of interest,

$$\sup |\hat{f}_m(\mathbf{t}, \gamma) - f(\mathbf{t}, \gamma)| \longrightarrow 0, \text{ almost surely,} \quad (23)$$

as $m \longrightarrow \infty$.

Let

$$I_1 = \{\mathbf{t} : f(\mathbf{t}, \gamma_0) < f(\hat{\mathbf{t}}, \gamma_0)\}, \quad \text{and} \quad I_2 = \{\mathbf{t} : \hat{f}_m(\mathbf{t}, \hat{\gamma}_T) < \hat{f}_m(\hat{\mathbf{t}}, \hat{\gamma}_T)\}.$$

Then

$$p_0 = \int_{I_1} f(\mathbf{t}, \gamma_0) d\mathbf{t}, \quad \text{and} \quad \hat{p}_{T,m,n} = \sum_{\mathbf{t}^{(i)} \in I_2} 1/n,$$

where $\mathbf{t}^{(i)}$, $i = 1, 2, \dots, n$, are n simulated values of \mathbf{t} from $f(\mathbf{t}, \hat{\gamma}_T)$. Note that as $n \longrightarrow \infty$,

$$\hat{p}_{T,m,n} \rightarrow \hat{p}_{T,m},$$

where

$$\hat{p}_{T,m} = \int_{I_2} f(\mathbf{t}, \hat{\gamma}_T) d\mathbf{t}.$$

Hence we need to show that $\hat{p}_{T,m} \rightarrow p_0$ as $T \rightarrow \infty$ and $m \rightarrow \infty$.

Note that I_2 can be rewritten as

$$I_2 = \{\mathbf{t} : \hat{f}_m(\hat{\mathbf{t}}, \hat{\gamma}_T) - f(\hat{\mathbf{t}}, \hat{\gamma}_T) + f(\hat{\mathbf{t}}, \hat{\gamma}_T) - f(\mathbf{t}, \hat{\gamma}_T) + f(\mathbf{t}, \hat{\gamma}_T) - \hat{f}_m(\mathbf{t}, \hat{\gamma}_T) > 0\}.$$

As $m \rightarrow \infty$,

$$I_2 \rightarrow I_3 = \{\mathbf{t} : f(\mathbf{t}, \hat{\gamma}_T) < f(\hat{\mathbf{t}}, \hat{\gamma}_T)\}, \text{ almost surely,}$$

because from [\(23\)](#),

$$\hat{f}_m(\hat{\mathbf{t}}, \hat{\gamma}_T) - f(\hat{\mathbf{t}}, \hat{\gamma}_T) \rightarrow 0, \text{ almost surely,}$$

and

$$f(\mathbf{t}, \hat{\gamma}_T) - \hat{f}_m(\mathbf{t}, \hat{\gamma}_T) \rightarrow 0, \text{ almost surely.}$$

Therefore, we now need to show that $\hat{p}_T \rightarrow p_0$ as $T \rightarrow \infty$, where

$$\hat{p}_T = \int_{I_3} f(\mathbf{t}, \hat{\gamma}_T) dt.$$

We have

$$\begin{aligned} p_0 - \hat{p}_T &= \int_{I_1 \cap I_3} (f(\mathbf{t}, \gamma_0) - f(\mathbf{t}, \hat{\gamma}_T)) dt + \int_{I_1 \cap \bar{I}_3} f(\mathbf{t}, \gamma_0) dt - \int_{\bar{I}_1 \cap I_3} f(\mathbf{t}, \hat{\gamma}_T) dt \\ &\stackrel{\Delta}{=} A_{1,T} + A_{2,T} - A_{3,T}. \end{aligned}$$

Let $I_s = \bigcup_T (I_1 \cap I_3)$ where $\bigcup_T (A)$ denotes the union for all values of T of the set A . It follows that

$$\begin{aligned} |A_{1,T}| &= \left| \int_{I_1 \cap I_3} (f(\mathbf{t}, \gamma_0) - f(\mathbf{t}, \hat{\gamma}_T)) dt \right| \\ &\leq \int_{I_1 \cap I_3} |f(\mathbf{t}, \gamma_0) - f(\mathbf{t}, \hat{\gamma}_T)| dt \\ &\leq \int_{I_s} |f(\mathbf{t}, \gamma_0) - f(\mathbf{t}, \hat{\gamma}_T)| dt \\ &\rightarrow 0, \text{ almost surely,} \end{aligned}$$

as $T \rightarrow \infty$ because of consistency of $\hat{\gamma}_T$ and because $f(\mathbf{t}, \gamma)$ is continuous in γ .

Note that $\mathbf{t} \in I_1 \cap \bar{I}_3$ implies

$$f(\mathbf{t}, \gamma_0) < f(\hat{\mathbf{t}}, \gamma_0), \text{ and } f(\mathbf{t}, \hat{\gamma}_T) \geq f(\hat{\mathbf{t}}, \hat{\gamma}_T).$$

Given that $\hat{\gamma}_T$ is consistent, $\hat{\gamma}_T$ almost surely converges to γ_0 . It suggests that $I_1 \cap \bar{I}_3$ has measure zero as $T \rightarrow \infty$. Therefore, $A_{2,T} \rightarrow 0$ as $T \rightarrow \infty$.

Similarly, we can show that $\bar{I}_1 \cap I_3$ has measure zero as $T \rightarrow \infty$, which implies $A_{3,T} \rightarrow 0$ as $T \rightarrow \infty$. This completes the proof. \square

Appendix 4: Proof of Theorem 3

Let

$$I_1 = \{t: f(t, \gamma_0) < f(\hat{t}, \gamma_0)\},$$

$$I_2 = \{t: \hat{f}_{T,m}^B(t) < \hat{f}_{T,m}^B(\hat{t})\}.$$

Then

$$p_0 = \int_{I_1} f(t, \gamma_0) dt \quad \text{and} \quad \hat{p}_{B,T,m,n} = \sum_{t^{(i)} \in I_2} 1/n,$$

where $t^{(i)}$, for $i = 1, 2, \dots, n$, are n bootstrapped values of t from $G_T(t, F_T)$. Observe that as $n \rightarrow \infty$,

$$\hat{p}_{B,T,m,n} \rightarrow \hat{p}_{B,T,m},$$

where

$$\hat{p}_{B,T,m} = \int_{I_2} dG_T(t, F_T).$$

Hence we need to show that $\hat{p}_{B,T,m} \rightarrow p_0$ as $T \rightarrow \infty$ and $m \rightarrow \infty$.

As $T \rightarrow \infty$, then

$$\begin{aligned} \hat{p}_{B,T,m} &\rightarrow \hat{p}_{B,\infty,m} = Pr \{t: \hat{f}_{\infty,m}^B(t) < \hat{f}_{\infty,m}^B(\hat{t}) \mid G_\infty(t, F_0)\} \\ &= Pr \{t: \hat{f}_m(t) < \hat{f}_m(\hat{t}) \mid G_\infty(t, F_0)\} \end{aligned}$$

because our bootstrap estimator is consistent and where $\hat{f}_m(t)$ denotes the kernel estimator of $f(t, \gamma_0)$ based on a random sample of size m from $G_\infty(t, F_0)$. Also, similarly, as $T \rightarrow \infty$,

$$I_2 \rightarrow I_2^* = \{t: \hat{f}_m(t) < \hat{f}_m(\hat{t}) \mid G_\infty(t, F_0)\},$$

so that

$$\hat{p}_{B,T,m} = \int_{I_2^*} dG_\infty(t, F_0) = \int_{I_2^*} f(t, \gamma_0) dt = \hat{p}_m.$$

Finally, $\hat{p}_m \rightarrow p_0$ as $m \rightarrow \infty$, because $\hat{f}_m(t) \rightarrow f(t, \gamma_0)$ and $I_2^* \rightarrow I_1$, which completes the proof.

□

References

- Baltagi, B.H., Bresson, G., Pirotte, A., 2006. Joint LM test for homoskedasticity in a one-way error component model. *Journal of Econometrics* 134, 401–417.
- Baltagi, B.H., Jung, B.C., Song, S.H., 2010. Testing for heteroskedasticity and serial correlation in a random effects panel data model. *Journal of Econometrics* 154, 122–124.
- Baltagi, B.H., Li, Q., 1995. Testing AR(1) against MA(1) disturbances in an error component model. *Journal of Econometrics* 68, 133–151.
- Barnard, G.A., 1963. Comment on ‘The spectral analysis of point processes’ by M. S. Bartlett. *Journal of the Royal Statistical Society, Series B* 25, 294.
- Beaulieu, M.C., Dufour, J.M., Khalaf, L., 2007. Multivariate tests of mean-variance efficiency with possibly non-Gaussian errors: An exact simulation-based approach. *Journal of Business & Economic Statistics* 25, 398–410.
- Beaulieu, M.C., Dufour, J.M., Khalaf, L., 2009. Finite sample multivariate tests of asset pricing models with coskewness. *Computational Statistics & Data Analysis* 53, 2008–2021.
- Bekaert, G., Ehrmann, M., Fratzscher, M., Mehl, A.J., 2014. Global crises and equity market contagion. *The Journal of Finance* 69, 2597–2649.
- Bera, A.K., Jarque, C.M., 1982. Model specification tests: A simultaneous approach. *Journal of Econometrics* 20, 59–82.
- Bernard, J.T., Idoudi, N., Khalaf, L., Yérou, C., 2007. Finite sample multivariate structural change tests with application to energy demand models. *Journal of Econometrics* 141, 1219–1244.
- Bowman, A.W., Azzalini, A., 1997. *Applied Smoothing Techniques for Data Analysis*. Oxford University Press, London.
- Bowman, K.O., Shenton, L.R., 1975. Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2 . *Biometrika* 62, 243–250.
- Box, G.E.P., Pierce, D.A., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509–1526.
- Breusch, T.S., Pagan, A.R., 1980. The Lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies* 47, 239–253.
- Breusch, T.S., Schmidt, P., 1988. Alternative forms of the Wald test: How long is a piece of string? *Communications in Statistics — Theory and Methods* 17, 2789–2795.
- Cadre, B., 2006. Kernel estimation of density level sets. *Journal of Multivariate Analysis* 97, 999–1023.
- Cadre, B., Pelletier, B., Pudlo, P., 2013. Estimation of density level sets with a given probability content. *Journal of Nonparametric Statistics* 25, 261–272.
- Chesher, A., 1983. The information matrix test: Simplified calculation via a score test interpretation. *Economics Letters* 13, 45–48.
- Chesher, A., 1984. Testing for neglected heterogeneity. *Econometrica* 52, 865–872.
- Chesher, A., Spady, R., 1991. Asymptotic expansions of the information matrix test statistic. *Econometrica* 59, 787–815.
- D’Agostino, R.B., 1971. An omnibus test of normality for moderate and large size samples. *Biometrika* 58, 341–348.
- D’Agostino, R.B., 1972. Small sample probability points for the D test of normality. *Biometrika* 59, 219–221.
- D’Agostino, R.B., Stephens, M.A., 1986. *Goodness-of-Fit Techniques*. Marcel Dekker, New York.
- Davidson, R., MacKinnon, J.G., 1992. A new form of the information matrix test. *Econometrica* 60, 145–157.
- Davidson, R., MacKinnon, J.G., 1993. *Estimation and Inference in Econometrics*. Oxford University Press, New York.

- Dhaene, G., Hoorelbeke, D., 2004. The information matrix test with bootstrap-based covariance matrix estimation. *Economics Letters* 82, 341–347.
- Dufour, J.M., 2006. Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics* 133, 443–477.
- Dufour, J.M., Farhat, A., Gardiol, L., Khalaf, L., 1998. Simulation-based finite sample normality tests in linear regressions. *The Econometrics Journal* 1, 154–173.
- Dufour, J.M., Farhat, A., Khalaf, L., 2004a. Tests multiples simulés et tests de normalité basés sur plusieurs moments dans les modèles de régression. *L'Actualité économique* 80, 501–522.
- Dufour, J.M., Khalaf, L., 1997. Bootstrap Methods in Econometrics: Theory and Numerical Performance, in: Kreps, D.M., Wilks, K.F. (Eds.), *Advances in Economics and Econometrics: Theory and Applications*. Cambridge University Press, Cambridge, U.K. volume 3, pp. 188–222.
- Dufour, J.M., Khalaf, L., 2001. Monte Carlo test methods in econometrics, in: Baltagi, B. (Ed.), *A Companion to Theoretical Econometrics*. Blackwell, Oxford, pp. 494–519.
- Dufour, J.M., Khalaf, L., 2002. Exact tests for contemporaneous correlation of disturbances in seemingly unrelated regressions. *Journal of Econometrics* 106, 143–170.
- Dufour, J.M., Khalaf, L., Beaulieu, M.C., 2003. Exact skewness-kurtosis tests for multivariate normality and goodness-of-fit in multivariate regressions with application to asset pricing models. *Oxford Bulletin of Economics and Statistics* 65, 891–906.
- Dufour, J.M., Khalaf, L., Beaulieu, M.C., 2010. Multivariate residual-based finite-sample tests for serial dependence and ARCH effects with applications to asset pricing models. *Journal of Applied Econometrics* 25, 263–285.
- Dufour, J.M., Khalaf, L., Bernard, J.T., Genest, I., 2004b. Simulation-based finite-sample tests for heteroskedasticity and ARCH effects. *Journal of Econometrics* 122, 317–347.
- Dufour, J.M., Khalaf, L., Voia, M., 2015. Finite-sample resampling-based combined hypothesis tests, with applications to serial correlation and predictability. *Communications in Statistics — Simulation and Computation* 44, 2329–2347.
- Dwass, M., 1957. Modified randomization tests for nonparametric hypotheses. *The Annals of Mathematical Statistics* 28, 181–187.
- Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.
- Engle, R.F., 1984. Wald, likelihood ratio, and Lagrange multiplier tests in econometrics, in: Griliches, Z., Intriligator, M.D. (Eds.), *Handbook of Econometrics*. North-Holland, Amsterdam. volume 2, pp. 775–826.
- Epanechnikov, V.A., 1969. Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications* 14, 153–158.
- Forbes, K.J., Rigobon, R., 2002. No contagion, only interdependence: Measuring stock market comovements. *The Journal of Finance* 57, 2223–2261.
- Godfrey, L.G., 1991. *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches*. 16, Cambridge University Press, Cambridge.
- Godfrey, L.G., 2005. Controlling the overall significance level of a battery of least squares diagnostic tests. *Oxford Bulletin of Economics & Statistics* 67, 263–279.
- Godfrey, L.G., 2009. *Bootstrap Tests for Regression Models*. Palgrave Macmillan, Basingstoke.
- Goh, K.L., King, M.L., 1999. A correction for local biasedness of the Wald and null Wald tests. *Oxford Bulletin of Economics and Statistics* 61, 435–450.
- Gregory, A.W., Veall, M.R., 1985. Formulating Wald tests of nonlinear restrictions. *Econometrica* , 1465–1468.
- Gregory, A.W., Veall, M.R., 1986. Wald tests of common factor restrictions. *Economics Letters* 22, 203–208.

- Gregory, A.W., Veall, M.R., 1987. Formulating Wald tests of the restrictions implied by the rational expectations hypothesis. *Journal of Applied Econometrics* 2, 61–68.
- Hall, A., 1987. The information matrix test for the linear model. *The Review of Economic Studies* 54, 257–263.
- Hall, P., 1992. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.
- Hall, P., 1994. Methodology and theory for the bootstrap, in: Engle, R.F., McFadden, D.F. (Eds.), *Handbook of Econometrics*. Elsevier Science B.V., Amsterdam. volume 4, pp. 2341–2381.
- Hayakawa, T., 1975. The likelihood ratio criterion for a composite hypothesis under a local alternative. *Biometrika* 62, 451–460.
- Holm, S., 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6, 65–70.
- Horowitz, J.L., 1994. Bootstrap-based critical values for the information matrix test. *Journal of Econometrics* 61, 395–411.
- Horowitz, J.L., 2001. The bootstrap, in: Heckman, J.L., Leamer, E. (Eds.), *Handbook of Econometrics*. Elsevier Science B.V., Amsterdam. volume 5, pp. 3159–3228.
- Horowitz, J.L., 2018. Bootstrap methods in econometrics. CEMMAP working paper, CWP53/18, Department of Economics, University College, London.
- Hyndman, R., 1996. Computing and graphing highest density regions. *The American Statistician* 50, 120–126.
- Jarque, C.M., Bera, A.K., 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* 6, 255–259.
- Jarque, C.M., Bera, A.K., 1987. A test for normality of observations and regression residuals. *International Statistical Review* 55, 163–172.
- King, M.L., 1981. The alternative Durbin-Watson test: An assessment of Durbin and Watson's choice of test statistic. *Journal of Econometrics* 17, 51–66.
- King, M.L., 1985. A point optimal test for autoregressive disturbances. *Journal of Econometrics* 27, 21–37.
- King, M.L., 1987. Testing for autocorrelation in linear regression models: A survey, in: King, M.L., Giles, D.E.A. (Eds.), *Specification Analysis in the Linear Model*. Routledge and Kegan Paul, London, pp. 19–73.
- King, M.L., Akram, M., Zhang, X., 2009. A new procedure for multiple one-sided hypothesis testing, in: *Econometric Society Australasian Meeting*, Canberra, July 2009.
- King, M.L., Goh, K.L., 2002. Improvements to the Wald test, in: Ullah, A., Wan, A.T.K., Chaturvedi, A. (Eds.), *Handbook of Applied Econometrics and Statistical Inference*. Marcel Dekker, New York, pp. 251–275.
- Lafontaine, F., White, K.J., 1986. Obtaining any Wald statistic you want. *Economics Letters* 21, 35–40.
- Lancaster, T., 1984. The covariance matrix of the information matrix test. *Econometrica* 52, 1051–1053.
- Lee, J.H.H., King, M.L., 1993. A locally most mean powerful based score test for ARCH and GARCH regression disturbances. *Journal of Business & Economic Statistics* 11, 17–27.
- Li, Q., Racine, J.S., 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, New Jersey.
- Ljung, G.M., Box, G.E.P., 1978. On a measure of lack of fit in time series models. *Biometrika* 65, 297–303.
- Longstaff, F.A., 2010. The subprime credit crisis and contagion in financial markets. *Journal of Financial Economics* 97, 436–450.
- MacKinnon, J.G., 2009. Bootstrap hypothesis testing, in: Belsley, D.A., Kontoghiorghes, E.J. (Eds.), *Handbook of Computational Econometrics*. John Wiley & Sons Ltd, Chichester, pp. 183–213.
- Mantel, N., 1987. Understanding Wald's test for exponential families. *The American Statistician* 41, 147–148.
- Mason, D.M., Polonik, W., 2009. Asymptotic normality of plug-in level set estimates. *The Annals of Applied Probability* 19, 1108–1142.

- Masry, E., 1996. Multivariate regression estimation local polynomial fitting for time series. *Stochastic Processes and Their Applications* 65, 81–101.
- Moulton, B.R., Randolph, W.C., 1989. Alternative tests of the error components model. *Econometrica* 57, 685–693.
- Orme, C., 1990. The small-sample performance of the information-matrix test. *Journal of Econometrics* 46, 309–331.
- Pagan, A., Vella, F., 1989. Diagnostic tests for models based on individual data: A survey. *Journal of Applied Econometrics* 4, S29–S59.
- Pearson, E.S., D’Agostino, R.B., Bowman, K.O., 1977. Tests for departure from normality: Comparison of powers. *Biometrika* 64, 231–246.
- Peers, H., 1971. Likelihood ratio and associated test criteria. *Biometrika* 58, 577–587.
- Poitras, G., 2006. More on the correct use of omnibus tests for normality. *Economics Letters* 90, 304–309.
- Racine, J.S., MacKinnon, J.G., 2007. Inference via kernel smoothing of bootstrap P values. *Computational Statistics & Data Analysis* 51, 5949–5957.
- Rapach, D.E., Strauss, J.K., Zhou, G., 2013. International stock return predictability: What is the role of the United States? *The Journal of Finance* 68, 1633–1662.
- Samworth, R.J., Wand, M.P., 2010. Asymptotics and optimal bandwidth selection for highest density region estimation. *The Annals of Statistics* 38, 1767–1792.
- Scott, D.W., 2015. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, New York.
- Seber, G.A.F., 1980. *The Linear Hypothesis: A General Theory*, Second Edition. Charles Griffin & Company Ltd, London.
- Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 591–611.
- Simes, R.J., 1986. An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 73, 751–754.
- Spiegelhalter, D.J., 1980. An omnibus test for normality for small samples. *Biometrika* 67, 493–496.
- Tauchen, G., 1985. Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* 30, 415–443.
- Taylor, L.W., 1987. The size bias of White’s information matrix test. *Economics Letters* 24, 63–67.
- Thode, H.C., 2002. *Testing for Normality*. Marcel Dekker, New York.
- Tsybakov, A.B., 1997. On nonparametric estimation of density level sets. *The Annals of Statistics* 25, 948–969.
- Urzúa, C.M., 1996. On the correct use of omnibus tests for normality. *Economics Letters* 53, 247–251.
- Wand, M.P., Jones, M.C., 1995. *Kernel Smoothing*. Chapman and Hall, New York.
- Watson, G.S., 1955. Serial correlation in regression analysis. I. *Biometrika* 42, 327–341.
- Westfall, P.H., Young, S.S., 1993. *Resampling-Based Multiple Testing: Examples and Methods for p -Value Adjustment*. John Wiley & Sons, New York.
- White, H., 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25.
- Wooldridge, J.M., 2001. Diagnostic testing, in: Baltagi, B. (Ed.), *A Companion to Theoretical Econometrics*. Blackwell, Oxford, pp. 180–200.
- Zhang, X., King, M.L., Hyndman, R.J., 2006. A Bayesian approach to bandwidth selection for multivariate kernel density estimation. *Computational Statistics & Data Analysis* 50, 3009–3031.

Table 1: The estimated sizes of the new test procedures, minimum p , product of p and Portman-teau tests when the regressor is $X1$, $d = 4$ and $d = 6$.

Dimension	T	α	Single simulation		Double simulation		Minimum	Product	Q
			NRR	MCMC	NRR	MCMC	p	p	
$d=4$	50	0.10	0.100	0.099	0.101	0.101	0.098	0.100	0.100
		0.05	0.051	0.051	0.048	0.048	0.050	0.051	0.052
		0.01	0.010	0.010	0.010	0.010	0.009	0.010	0.010
	100	0.10	0.098	0.101	0.101	0.101	0.103	0.101	0.102
		0.05	0.048	0.049	0.047	0.049	0.051	0.051	0.051
		0.01	0.009	0.009	0.009	0.009	0.010	0.010	0.011
	200	0.10	0.101	0.102	0.103	0.105	0.103	0.104	0.102
		0.05	0.050	0.050	0.050	0.051	0.053	0.051	0.052
		0.01	0.010	0.011	0.011	0.011	0.012	0.010	0.010
	500	0.10	0.097	0.095	0.093	0.096	0.097	0.095	0.095
		0.05	0.045	0.046	0.045	0.045	0.046	0.045	0.045
		0.01	0.009	0.009	0.009	0.009	0.010	0.009	0.009
$d=6$	50	0.10	0.104	0.103	0.102	0.102	0.100	0.104	0.104
		0.05	0.054	0.054	0.053	0.052	0.052	0.052	0.053
		0.01	0.010	0.010	0.010	0.010	0.010	0.011	0.010
	100	0.10	0.099	0.100	0.100	0.101	0.102	0.103	0.103
		0.05	0.049	0.050	0.050	0.050	0.051	0.049	0.050
		0.01	0.010	0.010	0.010	0.011	0.011	0.011	0.011
	200	0.10	0.101	0.100	0.102	0.102	0.102	0.102	0.103
		0.05	0.052	0.052	0.053	0.053	0.053	0.052	0.051
		0.01	0.012	0.011	0.010	0.011	0.012	0.009	0.010
	500	0.10	0.093	0.092	0.092	0.090	0.091	0.095	0.094
		0.05	0.044	0.045	0.045	0.046	0.046	0.044	0.046
		0.01	0.008	0.008	0.009	0.009	0.011	0.008	0.009

Table 2: Powers of the new tests and three competing tests estimated through parametric sampling when $d = 4$ and $\alpha = 0.05$, and the regressor is $X1$.

DGP	T	Single simulation		Double simulation		Minimum	Product	Q
		NRR	MCMC	NRR	MCMC	p	p	
AR(1): $\rho_1=0.25$	50	0.294	0.299	0.298	0.309	0.278	0.259	0.263
	100	0.520	0.529	0.519	0.531	0.549	0.482	0.504
	200	0.835	0.838	0.838	0.847	0.868	0.805	0.833
	500	0.997	0.998	0.997	0.989	0.999	0.996	0.998
AR(2): $\rho_1=0.05$ $\rho_2=0.10$	50	0.125	0.126	0.125	0.126	0.088	0.105	0.104
	100	0.163	0.166	0.167	0.169	0.132	0.149	0.148
	200	0.249	0.255	0.261	0.265	0.221	0.243	0.246
	500	0.521	0.531	0.523	0.531	0.489	0.507	0.519
AR(2): $\rho_1=0.05$ $\rho_2=0.20$	50	0.261	0.263	0.266	0.273	0.199	0.221	0.222
	100	0.419	0.424	0.418	0.429	0.373	0.378	0.387
	200	0.692	0.701	0.698	0.704	0.683	0.650	0.672
	500	0.978	0.979	0.978	0.979	0.981	0.969	0.976
AR(3): $\rho_1=0.10$ $\rho_2=0.10$ $\rho_3=0.10$	50	0.241	0.249	0.251	0.254	0.164	0.217	0.212
	100	0.371	0.374	0.372	0.380	0.291	0.369	0.365
	200	0.599	0.604	0.599	0.611	0.513	0.620	0.614
	500	0.933	0.931	0.934	0.939	0.881	0.946	0.943
AR(3): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.20$	50	0.271	0.271	0.269	0.270	0.190	0.214	0.219
	100	0.449	0.451	0.438	0.442	0.385	0.388	0.401
	200	0.480	0.500	0.490	0.501	0.488	0.449	0.481
	500	0.983	0.984	0.982	0.983	0.985	0.978	0.982
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.15$ $\rho_3=0.15$	50	0.422	0.424	0.424	0.429	0.283	0.367	0.367
	100	0.649	0.656	0.648	0.650	0.510	0.610	0.607
	200	0.890	0.895	0.890	0.894	0.806	0.880	0.878
	500	0.998	0.999	0.999	0.999	0.995	0.998	0.998
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.05$ $\rho_3=0.10$	50	0.236	0.237	0.238	0.242	0.149	0.198	0.198
	100	0.343	0.350	0.343	0.349	0.252	0.319	0.317
	200	0.542	0.551	0.548	0.556	0.438	0.539	0.535
	500	0.889	0.896	0.890	0.896	0.811	0.895	0.893
AR(4): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.05$ $\rho_3=0.05$	50	0.137	0.138	0.138	0.139	0.085	0.110	0.108
	100	0.172	0.175	0.168	0.171	0.124	0.155	0.152
	200	0.249	0.253	0.246	0.252	0.189	0.247	0.243
	500	0.483	0.496	0.485	0.496	0.377	0.504	0.496

Table 3: Sizes and powers of the new tests and three competing tests estimated through parametric sampling when $d = 4$ and the regressors are either $X2$ or $X3$.

DGP	T	α	$X2$					$X3$				
			NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p
Size	50	0.01	0.010	0.011	0.012	0.011	0.011	0.010	0.010	0.009	0.008	0.009
		0.05	0.050	0.050	0.050	0.049	0.050	0.048	0.049	0.047	0.046	0.047
		0.10	0.096	0.097	0.096	0.094	0.099	0.092	0.092	0.095	0.098	0.096
	100	0.01	0.011	0.011	0.010	0.009	0.010	0.011	0.011	0.010	0.011	0.010
		0.05	0.051	0.051	0.050	0.049	0.051	0.048	0.049	0.051	0.048	0.051
		0.10	0.102	0.102	0.101	0.102	0.102	0.107	0.106	0.103	0.099	0.102
	200	0.01	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.009	0.009	0.009
		0.05	0.050	0.049	0.050	0.051	0.047	0.049	0.048	0.050	0.049	0.051
		0.10	0.101	0.101	0.104	0.101	0.103	0.100	0.101	0.101	0.098	0.101
	500	0.01	0.008	0.008	0.010	0.010	0.010	0.010	0.009	0.010	0.011	0.010
		0.05	0.050	0.050	0.050	0.049	0.049	0.051	0.051	0.049	0.048	0.049
		0.10	0.100	0.100	0.101	0.101	0.102	0.099	0.099	0.101	0.100	0.101
Power												
AR(1): $\rho_1=0.25$	50	0.05	0.203	0.207	0.116	0.069	0.121	0.175	0.178	0.123	0.082	0.129
	100	0.05	0.447	0.455	0.250	0.222	0.252	0.403	0.411	0.320	0.318	0.315
	200	0.05	0.781	0.791	0.653	0.695	0.621	0.768	0.777	0.733	0.775	0.704
	500	0.05	0.996	0.997	0.994	0.997	0.991	0.996	0.997	0.996	0.998	0.994
AR(2): $\rho_1=0.05$ $\rho_2=0.10$	50	0.05	0.087	0.089	0.055	0.050	0.054	0.086	0.088	0.039	0.035	0.039
	100	0.05	0.131	0.131	0.057	0.048	0.058	0.131	0.134	0.065	0.055	0.067
	200	0.05	0.214	0.216	0.104	0.093	0.104	0.217	0.218	0.139	0.119	0.139
	500	0.05	0.521	0.527	0.375	0.355	0.364	0.525	0.534	0.427	0.394	0.417
AR(2): $\rho_1=0.05$ $\rho_2=0.20$	50	0.05	0.165	0.166	0.094	0.071	0.096	0.166	0.168	0.045	0.041	0.044
	100	0.05	0.335	0.336	0.144	0.115	0.144	0.340	0.349	0.158	0.146	0.158
	200	0.05	0.617	0.623	0.407	0.421	0.387	0.627	0.634	0.474	0.481	0.457
	500	0.05	0.974	0.977	0.946	0.955	0.931	0.974	0.978	0.956	0.962	0.944
AR(3): $\rho_1=0.10$ $\rho_2=0.10$ $\rho_3=0.10$	50	0.05	0.104	0.105	0.036	0.033	0.036	0.111	0.112	0.036	0.033	0.035
	100	0.05	0.227	0.228	0.056	0.048	0.059	0.231	0.237	0.109	0.088	0.114
	200	0.05	0.474	0.479	0.268	0.199	0.271	0.475	0.484	0.383	0.291	0.392
	500	0.05	0.917	0.919	0.855	0.753	0.860	0.919	0.923	0.897	0.812	0.901
AR(3): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.20$	50	0.05	0.128	0.127	0.081	0.051	0.089	0.121	0.122	0.081	0.050	0.084
	100	0.05	0.294	0.292	0.117	0.094	0.117	0.271	0.277	0.168	0.150	0.166
	200	0.05	0.608	0.615	0.394	0.408	0.371	0.596	0.605	0.505	0.509	0.482
	500	0.05	0.974	0.975	0.953	0.959	0.942	0.973	0.974	0.968	0.973	0.960
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.15$ $\rho_4=0.15$	50	0.05	0.118	0.117	0.060	0.055	0.064	0.172	0.173	0.042	0.035	0.042
	100	0.05	0.332	0.329	0.083	0.062	0.086	0.408	0.416	0.179	0.133	0.183
	200	0.05	0.733	0.738	0.483	0.380	0.485	0.785	0.790	0.653	0.534	0.656
	500	0.05	0.995	0.996	0.989	0.970	0.989	0.997	0.997	0.994	0.983	0.994
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.05$ $\rho_4=0.10$	50	0.05	0.098	0.097	0.056	0.049	0.058	0.112	0.113	0.027	0.025	0.027
	100	0.05	0.195	0.194	0.056	0.043	0.058	0.221	0.227	0.074	0.061	0.076
	200	0.05	0.396	0.400	0.192	0.149	0.194	0.428	0.430	0.281	0.208	0.287
	500	0.05	0.858	0.861	0.748	0.643	0.751	0.865	0.871	0.804	0.697	0.812
AR(4): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.05$ $\rho_4=0.05$	50	0.05	0.071	0.069	0.041	0.039	0.042	0.071	0.072	0.033	0.032	0.033
	100	0.05	0.106	0.104	0.032	0.028	0.034	0.108	0.111	0.042	0.039	0.043
	200	0.05	0.176	0.178	0.061	0.052	0.061	0.184	0.187	0.105	0.083	0.108
	500	0.05	0.441	0.445	0.295	0.219	0.302	0.450	0.456	0.368	0.274	0.376

Table 4: Sizes of the new tests and three competing tests estimated via bootstrap sampling when $d = 4$ and the error term is either normal or mixture normal.

Regressors	T	α	Normal errors					Mixture normal errors				
			NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p
X1	50	0.01	0.007	0.008	0.009	0.008	0.009	0.010	0.010	0.011	0.013	0.012
		0.05	0.044	0.043	0.044	0.046	0.044	0.049	0.048	0.047	0.051	0.047
		0.10	0.090	0.090	0.093	0.090	0.091	0.095	0.094	0.096	0.099	0.094
	100	0.01	0.010	0.010	0.011	0.011	0.011	0.010	0.010	0.010	0.009	0.010
		0.05	0.046	0.046	0.047	0.046	0.049	0.047	0.048	0.048	0.049	0.049
		0.10	0.097	0.098	0.097	0.098	0.097	0.099	0.098	0.100	0.099	0.102
	200	0.01	0.011	0.011	0.011	0.011	0.011	0.012	0.012	0.010	0.011	0.011
		0.05	0.050	0.049	0.049	0.051	0.048	0.053	0.053	0.051	0.052	0.051
		0.10	0.101	0.101	0.098	0.100	0.101	0.104	0.104	0.099	0.097	0.098
500	0.01	0.010	0.010	0.011	0.011	0.011	0.009	0.009	0.009	0.009	0.009	
	0.05	0.049	0.049	0.050	0.049	0.050	0.052	0.051	0.046	0.047	0.046	
	0.10	0.099	0.100	0.099	0.099	0.099	0.100	0.099	0.092	0.093	0.094	
X2	50	0.01	0.009	0.009	0.007	0.009	0.007	0.009	0.010	0.009	0.010	0.009
		0.05	0.047	0.047	0.039	0.048	0.041	0.048	0.048	0.046	0.048	0.046
		0.10	0.101	0.101	0.086	0.092	0.082	0.095	0.094	0.095	0.095	0.095
	100	0.01	0.007	0.007	0.009	0.008	0.008	0.020	0.021	0.020	0.015	0.020
		0.05	0.048	0.051	0.046	0.047	0.046	0.081	0.082	0.068	0.063	0.070
		0.10	0.116	0.114	0.095	0.105	0.095	0.145	0.147	0.127	0.120	0.127
	200	0.01	0.009	0.009	0.013	0.014	0.011	0.011	0.011	0.011	0.009	0.011
		0.05	0.051	0.051	0.054	0.057	0.054	0.052	0.050	0.050	0.050	0.050
		0.10	0.099	0.100	0.106	0.102	0.105	0.103	0.102	0.101	0.101	0.100
500	0.01	0.008	0.009	0.008	0.005	0.008	0.011	0.010	0.010	0.009	0.011	
	0.05	0.054	0.054	0.048	0.047	0.048	0.048	0.048	0.051	0.050	0.050	
	0.10	0.102	0.104	0.099	0.101	0.098	0.097	0.098	0.099	0.098	0.100	
X3	50	0.01	0.010	0.010	0.011	0.010	0.011	0.009	0.009	0.009	0.011	0.009
		0.05	0.049	0.049	0.052	0.051	0.052	0.048	0.048	0.047	0.047	0.047
		0.10	0.105	0.104	0.101	0.105	0.104	0.095	0.095	0.097	0.094	0.097
	100	0.01	0.008	0.008	0.009	0.010	0.009	0.010	0.010	0.011	0.012	0.011
		0.05	0.050	0.050	0.048	0.047	0.048	0.049	0.049	0.049	0.052	0.048
		0.10	0.095	0.097	0.095	0.094	0.096	0.098	0.098	0.099	0.102	0.099
	200	0.01	0.009	0.009	0.010	0.011	0.010	0.007	0.007	0.007	0.008	0.007
		0.05	0.050	0.051	0.050	0.049	0.050	0.042	0.041	0.044	0.044	0.044
		0.10	0.102	0.101	0.101	0.097	0.102	0.091	0.092	0.093	0.093	0.092
500	0.01	0.009	0.010	0.009	0.009	0.009	0.009	0.009	0.008	0.009	0.007	
	0.05	0.049	0.050	0.049	0.050	0.048	0.046	0.047	0.047	0.050	0.047	
	0.10	0.098	0.097	0.097	0.097	0.098	0.099	0.099	0.098	0.095	0.097	

Table 5: Powers of the new tests and three competing tests estimated via bootstrap sampling when $d = 4$, $\alpha = 0.05$, the error term is either normal or mixture normal, and the regressor is $X1$.

DGP	T	Normal errors					Mixture normal errors				
		NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p
AR(1): $\rho_1=0.25$	50	0.191	0.198	0.194	0.209	0.189	0.214	0.222	0.215	0.219	0.214
	100	0.460	0.472	0.439	0.481	0.422	0.302	0.304	0.252	0.365	0.228
	200	0.798	0.804	0.800	0.846	0.772	0.807	0.818	0.801	0.848	0.772
	500	0.997	0.998	0.997	0.999	0.996	0.998	0.998	0.998	0.999	0.996
AR(2): $\rho_1=0.05$ $\rho_2=0.10$	50	0.077	0.078	0.071	0.065	0.070	0.082	0.083	0.078	0.071	0.079
	100	0.129	0.132	0.115	0.102	0.118	0.114	0.116	0.114	0.106	0.115
	200	0.197	0.198	0.204	0.193	0.202	0.211	0.217	0.210	0.192	0.207
	500	0.486	0.496	0.507	0.473	0.495	0.488	0.495	0.497	0.470	0.488
AR(2): $\rho_1=0.05$ $\rho_2=0.20$	50	0.166	0.170	0.152	0.143	0.150	0.161	0.164	0.161	0.148	0.161
	100	0.340	0.349	0.317	0.308	0.311	0.321	0.327	0.331	0.322	0.322
	200	0.611	0.617	0.619	0.642	0.597	0.653	0.657	0.628	0.646	0.605
	500	0.972	0.973	0.973	0.978	0.966	0.957	0.958	0.967	0.975	0.959
AR(3): $\rho_1=0.10$ $\rho_2=0.10$ $\rho_3=0.10$	50	0.122	0.126	0.123	0.103	0.124	0.069	0.070	0.095	0.089	0.096
	100	0.293	0.299	0.274	0.220	0.279	0.282	0.287	0.277	0.228	0.279
	200	0.501	0.506	0.545	0.455	0.551	0.516	0.525	0.549	0.447	0.556
	500	0.922	0.928	0.935	0.867	0.938	0.922	0.925	0.934	0.868	0.937
AR(3): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.20$	50	0.147	0.149	0.135	0.126	0.132	0.172	0.171	0.147	0.141	0.144
	100	0.362	0.367	0.312	0.303	0.303	0.314	0.318	0.309	0.306	0.299
	200	0.624	0.632	0.644	0.658	0.623	0.603	0.613	0.635	0.642	0.614
	500	0.981	0.981	0.980	0.983	0.976	0.967	0.971	0.978	0.980	0.973
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.15$ $\rho_4=0.15$	50	0.219	0.222	0.211	0.163	0.210	0.261	0.265	0.222	0.171	0.223
	100	0.525	0.531	0.479	0.387	0.485	0.506	0.516	0.492	0.398	0.495
	200	0.797	0.803	0.825	0.748	0.827	0.816	0.819	0.827	0.743	0.830
	500	0.997	0.997	0.998	0.993	0.998	0.990	0.991	0.996	0.989	0.997
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.05$ $\rho_4=0.10$	50	0.117	0.119	0.115	0.093	0.115	0.148	0.149	0.123	0.100	0.124
	100	0.247	0.254	0.229	0.182	0.235	0.263	0.266	0.238	0.192	0.240
	200	0.411	0.415	0.461	0.379	0.466	0.483	0.488	0.468	0.383	0.471
	500	0.855	0.864	0.877	0.789	0.881	0.866	0.872	0.872	0.787	0.875
AR(4): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.05$ $\rho_4=0.05$	50	0.068	0.069	0.063	0.057	0.063	0.079	0.079	0.071	0.065	0.072
	100	0.121	0.123	0.107	0.089	0.109	0.100	0.104	0.102	0.085	0.104
	200	0.167	0.168	0.194	0.155	0.198	0.148	0.150	0.194	0.153	0.197
	500	0.437	0.447	0.470	0.354	0.479	0.373	0.380	0.459	0.353	0.468

Table 6: Powers of the new tests and three competing tests estimated via bootstrap sampling when $d = 4$, $\alpha = 0.05$, the error term is either normal or mixture normal, and the regressors are X_2 .

DGP	T	Normal errors					Mixture normal errors				
		NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p
AR(1): $\rho_1=0.25$	50	0.331	0.336	0.083	0.048	0.088	0.341	0.345	0.086	0.051	0.092
	100	0.541	0.551	0.249	0.248	0.246	0.541	0.548	0.239	0.232	0.233
	200	0.835	0.848	0.698	0.750	0.666	0.822	0.828	0.682	0.731	0.643
	500	0.998	0.998	0.996	0.998	0.993	0.997	0.998	0.996	0.998	0.993
AR(2): $\rho_1=0.05$ $\rho_2=0.10$	50	0.188	0.189	0.023	0.019	0.022	0.186	0.186	0.024	0.022	0.024
	100	0.197	0.200	0.041	0.033	0.041	0.180	0.182	0.036	0.031	0.037
	200	0.258	0.266	0.117	0.102	0.117	0.233	0.236	0.105	0.094	0.104
	500	0.523	0.532	0.409	0.375	0.398	0.524	0.530	0.409	0.376	0.400
AR(2): $\rho_1=0.05$ $\rho_2=0.20$	50	0.312	0.312	0.025	0.023	0.024	0.306	0.307	0.025	0.024	0.025
	100	0.424	0.427	0.112	0.103	0.112	0.402	0.409	0.100	0.089	0.098
	200	0.661	0.671	0.435	0.447	0.419	0.643	0.648	0.417	0.429	0.394
	500	0.973	0.976	0.952	0.959	0.940	0.976	0.976	0.953	0.959	0.941
AR(3): $\rho_1=0.10$ $\rho_2=0.10$ $\rho_3=0.10$	50	0.273	0.274	0.020	0.018	0.020	0.270	0.270	0.023	0.020	0.023
	100	0.374	0.378	0.076	0.057	0.079	0.345	0.350	0.070	0.052	0.073
	200	0.583	0.593	0.351	0.263	0.359	0.562	0.569	0.332	0.244	0.336
	500	0.930	0.935	0.891	0.797	0.895	0.927	0.933	0.885	0.793	0.889
AR(3): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.20$	50	0.294	0.293	0.050	0.029	0.054	0.289	0.287	0.058	0.034	0.063
	100	0.446	0.448	0.117	0.104	0.116	0.402	0.404	0.115	0.095	0.113
	200	0.698	0.705	0.462	0.476	0.442	0.695	0.699	0.446	0.457	0.417
	500	0.981	0.982	0.965	0.970	0.957	0.982	0.984	0.963	0.968	0.954
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.15$ $\rho_4=0.15$	50	0.389	0.384	0.025	0.019	0.025	0.369	0.365	0.027	0.025	0.029
	100	0.549	0.551	0.133	0.090	0.137	0.506	0.509	0.121	0.083	0.125
	200	0.845	0.848	0.620	0.499	0.623	0.825	0.829	0.597	0.477	0.598
	500	0.997	0.998	0.993	0.981	0.994	0.997	0.998	0.993	0.979	0.994
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.05$ $\rho_4=0.10$	50	0.268	0.266	0.014	0.012	0.015	0.256	0.253	0.017	0.017	0.017
	100	0.313	0.317	0.050	0.038	0.053	0.282	0.287	0.046	0.034	0.047
	200	0.501	0.508	0.251	0.185	0.259	0.466	0.471	0.237	0.173	0.238
	500	0.878	0.885	0.793	0.679	0.801	0.868	0.876	0.787	0.675	0.792
AR(4): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.05$ $\rho_4=0.05$	50	0.189	0.189	0.019	0.017	0.019	0.182	0.182	0.021	0.021	0.021
	100	0.191	0.193	0.025	0.022	0.025	0.169	0.172	0.024	0.022	0.026
	200	0.248	0.253	0.087	0.071	0.091	0.228	0.230	0.082	0.065	0.083
	500	0.483	0.493	0.354	0.257	0.362	0.476	0.488	0.354	0.255	0.360

Table 7: Powers of the new tests and three competing tests estimated via bootstrap sampling when $d = 4$, $\alpha = 0.05$, the error term is either normal or mixture normal, and the regressors are X_3 .

DGP	T	Normal errors					Mixture normal errors				
		NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p
AR(1): $\rho_1=0.25$	50	0.172	0.174	0.122	0.080	0.131	0.166	0.169	0.124	0.088	0.131
	100	0.401	0.408	0.323	0.318	0.313	0.379	0.388	0.309	0.303	0.300
	200	0.770	0.782	0.725	0.772	0.694	0.775	0.783	0.725	0.771	0.696
	500	0.995	0.996	0.996	0.998	0.994	0.997	0.997	0.997	0.998	0.995
AR(2): $\rho_1=0.05$ $\rho_2=0.10$	50	0.091	0.091	0.039	0.037	0.040	0.084	0.085	0.040	0.039	0.039
	100	0.124	0.125	0.063	0.058	0.064	0.122	0.125	0.062	0.058	0.061
	200	0.214	0.215	0.137	0.121	0.137	0.209	0.216	0.131	0.117	0.131
	500	0.497	0.503	0.425	0.399	0.417	0.487	0.488	0.425	0.396	0.418
AR(2): $\rho_1=0.05$ $\rho_2=0.20$	50	0.153	0.154	0.039	0.039	0.039	0.191	0.191	0.042	0.042	0.041
	100	0.322	0.324	0.160	0.154	0.158	0.315	0.322	0.148	0.139	0.146
	200	0.625	0.631	0.472	0.488	0.451	0.617	0.626	0.468	0.470	0.450
	500	0.971	0.974	0.959	0.965	0.950	0.970	0.973	0.956	0.962	0.945
AR(3): $\rho_1=0.10$ $\rho_2=0.10$ $\rho_3=0.10$	50	0.118	0.117	0.039	0.036	0.039	0.115	0.114	0.035	0.035	0.035
	100	0.226	0.229	0.113	0.092	0.115	0.213	0.220	0.100	0.082	0.103
	200	0.490	0.492	0.379	0.292	0.388	0.484	0.487	0.364	0.282	0.370
	500	0.911	0.915	0.896	0.808	0.901	0.919	0.923	0.898	0.813	0.903
AR(3): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.20$	50	0.123	0.123	0.086	0.057	0.086	0.127	0.128	0.091	0.061	0.094
	100	0.274	0.277	0.168	0.161	0.168	0.258	0.264	0.161	0.149	0.156
	200	0.601	0.609	0.501	0.525	0.478	0.605	0.612	0.504	0.512	0.479
	500	0.975	0.977	0.968	0.972	0.959	0.974	0.976	0.969	0.973	0.963
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.15$ $\rho_4=0.15$	50	0.174	0.174	0.044	0.043	0.043	0.175	0.175	0.045	0.039	0.045
	100	0.411	0.416	0.177	0.136	0.178	0.399	0.400	0.169	0.128	0.173
	200	0.788	0.794	0.650	0.540	0.654	0.780	0.785	0.652	0.540	0.656
	500	0.997	0.997	0.995	0.985	0.995	0.998	0.998	0.995	0.984	0.995
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.05$ $\rho_4=0.10$	50	0.115	0.114	0.028	0.030	0.028	0.111	0.112	0.028	0.028	0.028
	100	0.212	0.215	0.072	0.062	0.073	0.201	0.206	0.072	0.058	0.073
	200	0.442	0.447	0.288	0.221	0.293	0.419	0.424	0.269	0.207	0.272
	500	0.861	0.865	0.811	0.700	0.813	0.849	0.857	0.803	0.698	0.806
AR(4): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.05$ $\rho_4=0.05$	50	0.083	0.082	0.036	0.035	0.036	0.083	0.082	0.036	0.035	0.036
	100	0.108	0.112	0.043	0.043	0.044	0.108	0.112	0.043	0.043	0.044
	200	0.180	0.183	0.105	0.089	0.107	0.180	0.183	0.105	0.089	0.107
	500	0.428	0.441	0.364	0.276	0.373	0.428	0.441	0.364	0.276	0.373

Table 8: Powers of the new tests estimated through parametric sampling when the regressor is X_1 and $d = 20$ and 50 .

Sample size	m	$d = 20$			$d = 50$		
		0.01	0.05	0.10	0.01	0.05	0.10
100	10,000	0.1240	0.2804	0.3943	0.1040	0.2472	0.3441
	20,000	0.1210	0.2734	0.3944	0.1069	0.2390	0.3362
	30,000	0.1167	0.2713	0.3838	0.1054	0.2380	0.3370
	50,000	0.1106	0.2688	0.3830	0.1029	0.2331	0.3376
	75,000	0.1118	0.2647	0.3818	0.1025	0.2332	0.3366
	100,000	0.1128	0.2652	0.3847	0.1053	0.2365	0.3355
200	10,000	0.2851	0.5143	0.6259	0.1645	0.3744	0.4887
	20,000	0.2826	0.5197	0.6367	0.1723	0.3667	0.4887
	30,000	0.2767	0.5073	0.6356	0.1655	0.3584	0.4828
	50,000	0.2820	0.5135	0.6397	0.1706	0.3631	0.4847
	75,000	0.2845	0.5161	0.6395	0.1766	0.3679	0.4909
	100,000	0.2885	0.5160	0.6440	0.1736	0.3664	0.4905
300	10,000	0.5021	0.7173	0.8154	0.2843	0.5205	0.6376
	20,000	0.5063	0.7202	0.8113	0.2873	0.5164	0.6411
	30,000	0.5016	0.7278	0.8192	0.2970	0.5219	0.6486
	50,000	0.5032	0.7238	0.8173	0.3003	0.5277	0.6510
	75,000	0.5054	0.7299	0.8211	0.3005	0.5318	0.6525
	100,000	0.5062	0.7291	0.8223	0.3011	0.5324	0.6535
500	10,000	0.8222	0.9301	0.9598	0.5868	0.7889	0.8684
	20,000	0.8278	0.9338	0.9661	0.5941	0.7918	0.8704
	30,000	0.8231	0.9335	0.9647	0.5881	0.7965	0.8717
	50,000	0.8309	0.9356	0.9656	0.5888	0.8012	0.8756
	75,000	0.8357	0.9378	0.9674	0.5989	0.8041	0.8772
	100,000	0.8385	0.9396	0.9677	0.6105	0.8069	0.8809
1000	10,000	0.9983	0.9997	0.9998	0.9701	0.9928	0.9965
	20,000	0.9985	0.9997	0.9998	0.9727	0.9938	0.9975
	30,000	0.9982	0.9997	0.9998	0.9729	0.9934	0.9973
	50,000	0.9985	0.9997	0.9999	0.9759	0.9943	0.9977
	75,000	0.9987	0.9998	0.9999	0.9753	0.9946	0.9977
	100,000	0.9986	0.9998	0.9999	0.9757	0.9944	0.9979

Table 9: Estimated sizes and powers of the new test procedures, Dufour's (2006) exact tests and JB test.

	T	α	Single simulation		Double simulation		Minimum	Product	Q
			NRR	MCMC	NRR	MCMC	p	p	
Size	30	0.10	0.099	0.099	0.100	0.100	0.101	0.099	0.099
		0.05	0.051	0.051	0.050	0.050	0.050	0.050	0.051
		0.01	0.010	0.010	0.010	0.010	0.011	0.011	0.011
	50	0.10	0.099	0.098	0.098	0.097	0.099	0.099	0.101
		0.05	0.046	0.046	0.047	0.048	0.048	0.048	0.048
		0.01	0.011	0.011	0.011	0.010	0.011	0.011	0.011
	75	0.10	0.104	0.104	0.103	0.103	0.102	0.103	0.104
		0.05	0.051	0.051	0.051	0.052	0.051	0.051	0.052
		0.01	0.010	0.010	0.009	0.009	0.010	0.010	0.011
	100	0.10	0.103	0.103	0.103	0.102	0.101	0.100	0.100
		0.05	0.053	0.053	0.053	0.053	0.054	0.055	0.054
		0.01	0.011	0.011	0.011	0.011	0.011	0.011	0.011
Student t_5	30	0.10	0.416	0.414	0.417	0.416	0.385	0.397	0.396
		0.05	0.324	0.322	0.320	0.319	0.299	0.310	0.309
		0.01	0.177	0.178	0.179	0.177	0.179	0.179	0.180
	50	0.10	0.542	0.544	0.540	0.539	0.502	0.519	0.519
		0.05	0.442	0.445	0.444	0.445	0.421	0.433	0.432
		0.01	0.289	0.290	0.284	0.283	0.287	0.278	0.280
	75	0.10	0.665	0.666	0.664	0.664	0.620	0.637	0.637
		0.05	0.568	0.569	0.569	0.569	0.547	0.558	0.556
		0.01	0.379	0.381	0.383	0.382	0.389	0.374	0.379
	100	0.10	0.743	0.743	0.738	0.739	0.704	0.717	0.718
		0.05	0.659	0.659	0.663	0.663	0.641	0.652	0.652
		0.01	0.480	0.480	0.478	0.480	0.489	0.469	0.475
χ_3^2	30	0.10	0.829	0.834	0.832	0.835	0.848	0.845	0.844
		0.05	0.730	0.740	0.738	0.744	0.745	0.647	0.683
		0.01	0.489	0.503	0.478	0.481	0.478	0.395	0.399
	50	0.10	0.976	0.976	0.977	0.977	0.970	0.977	0.978
		0.05	0.948	0.945	0.947	0.947	0.930	0.880	0.900
		0.01	0.818	0.809	0.804	0.805	0.759	0.635	0.647
	75	0.10	0.999	0.666	0.664	0.664	0.620	0.637	0.637
		0.05	0.997	0.997	0.996	0.996	0.992	0.990	0.991
		0.01	0.973	0.971	0.976	0.976	0.922	0.823	0.841
	100	0.10	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.05	1.000	1.000	1.000	1.000	0.999	1.000	1.000
		0.01	0.998	0.998	0.998	0.998	0.986	0.941	0.956

Table 9: Estimated sizes and powers of the new test procedures, Dufour's (2006) exact tests and JB test (continued).

	T	α	Single simulation		Double simulation		Minimum	Product	Q
			NRR	MCMC	NRR	MCMC	p	p	
Gamma	30	0.10	0.720	0.723	0.724	0.726	0.764	0.745	0.749
		0.05	0.601	0.608	0.612	0.616	0.639	0.545	0.576
		0.01	0.359	0.370	0.349	0.350	0.366	0.302	0.305
	50	0.10	0.933	0.933	0.935	0.935	0.930	0.936	0.938
		0.05	0.874	0.869	0.876	0.877	0.864	0.786	0.815
		0.01	0.670	0.660	0.654	0.656	0.632	0.514	0.523
	75	0.10	0.993	0.993	0.994	0.994	0.990	0.994	0.994
		0.05	0.981	0.981	0.981	0.981	0.972	0.954	0.962
		0.01	0.905	0.902	0.909	0.910	0.829	0.702	0.720
100	0.10	1.000	1.000	1.000	1.000	0.999	1.000	1.000	
	0.05	0.998	0.998	0.998	0.998	0.995	0.994	0.995	
	0.01	0.980	0.979	0.980	0.979	0.949	0.859	0.881	
log normal	30	0.10	0.718	0.720	0.722	0.723	0.773	0.746	0.754
		0.05	0.616	0.619	0.625	0.628	0.670	0.598	0.622
		0.01	0.420	0.430	0.407	0.407	0.436	0.385	0.388
	50	0.10	0.920	0.921	0.923	0.923	0.934	0.927	0.932
		0.05	0.869	0.866	0.873	0.874	0.880	0.822	0.843
		0.01	0.704	0.699	0.696	0.697	0.694	0.609	0.614
	75	0.10	0.988	0.988	0.990	0.990	0.989	0.990	0.990
		0.05	0.975	0.974	0.975	0.975	0.975	0.956	0.963
		0.01	0.905	0.903	0.907	0.907	0.869	0.787	0.798
	100	0.10	0.999	0.999	0.998	0.998	0.998	0.999	0.999
		0.05	0.996	0.996	0.996	0.996	0.995	0.993	0.994
		0.01	0.977	0.977	0.976	0.975	0.961	0.909	0.921

Table 10: Estimated sizes of the IM tests with samples generated from the linear regression model.

Test	Sample size	One regressor			Two regressors		
		0.01	0.05	0.10	0.01	0.05	0.10
New test	50	0.013	0.047	0.095	0.011	0.052	0.108
	100	0.010	0.050	0.097	0.008	0.047	0.097
	200	0.009	0.047	0.098	0.010	0.054	0.104
	300	0.007	0.047	0.098	0.011	0.046	0.092
IM _L	50	0.013	0.053	0.106	0.008	0.052	0.102
	100	0.010	0.049	0.101	0.009	0.047	0.098
	200	0.009	0.056	0.102	0.010	0.051	0.105
	300	0.007	0.054	0.104	0.009	0.050	0.104
IM _{DH}	50	0.011	0.049	0.090	0.012	0.051	0.110
	100	0.010	0.048	0.098	0.007	0.046	0.092
	200	0.009	0.047	0.098	0.010	0.052	0.103
	300	0.010	0.046	0.096	0.008	0.044	0.096
minp _L	50	0.012	0.053	0.101	0.008	0.045	0.096
	100	0.010	0.051	0.096	0.010	0.046	0.098
	200	0.011	0.051	0.102	0.012	0.057	0.108
	300	0.010	0.050	0.100	0.009	0.047	0.098
prod _L	50	0.011	0.049	0.090	0.012	0.051	0.110
	100	0.010	0.048	0.098	0.007	0.046	0.092
	200	0.009	0.047	0.098	0.010	0.052	0.103
	300	0.010	0.046	0.096	0.008	0.044	0.096
minp _{DH}	50	0.007	0.048	0.095	0.011	0.050	0.107
	100	0.011	0.049	0.096	0.010	0.049	0.096
	200	0.010	0.050	0.099	0.011	0.051	0.097
	300	0.009	0.049	0.098	0.009	0.046	0.095
prod _{DH}	50	0.008	0.046	0.097	0.013	0.055	0.108
	100	0.011	0.048	0.098	0.007	0.046	0.095
	200	0.008	0.046	0.103	0.012	0.051	0.100
	300	0.010	0.050	0.101	0.009	0.046	0.102

Table 11: Estimated powers of the IM tests with samples generated from the linear regression model.

Test	Sample size	One regressor			Two regressors		
		0.01	0.05	0.10	0.01	0.05	0.10
New test	50	0.362	0.639	0.746	0.239	0.456	0.596
	100	0.010	0.728	0.910	0.648	0.869	0.930
	200	0.983	0.999	0.999	0.966	0.994	0.998
	300	0.999	1.000	1.000	0.998	1.000	1.000
IM _L	50	0.006	0.038	0.090	0.005	0.030	0.064
	100	0.022	0.198	0.414	0.012	0.081	0.182
	200	0.375	0.749	0.886	0.144	0.437	0.624
	300	0.759	0.954	0.990	0.357	0.728	0.871
IM _{DH}	50	0.153	0.329	0.442	0.184	0.376	0.484
	100	0.253	0.537	0.679	0.304	0.533	0.644
	200	0.655	0.887	0.940	0.488	0.744	0.826
	300	0.884	0.973	0.988	0.699	0.887	0.938
minp _L	50	0.013	0.196	0.396	0.003	0.054	0.172
	100	0.283	0.636	0.790	0.064	0.338	0.508
	200	0.757	0.944	0.981	0.388	0.776	0.899
	300	0.921	0.992	0.998	0.679	0.934	0.982
prod _L	50	0.102	0.279	0.412	0.090	0.210	0.310
	100	0.315	0.683	0.821	0.301	0.570	0.703
	200	0.858	0.974	0.992	0.806	0.951	0.981
	300	0.981	0.998	1.000	0.976	0.997	0.999
minp _{DH}	50	0.148	0.276	0.361	0.163	0.287	0.367
	100	0.311	0.538	0.646	0.247	0.444	0.558
	200	0.673	0.867	0.925	0.497	0.736	0.830
	300	0.879	0.967	0.986	0.715	0.898	0.946
prod _{DH}	50	0.135	0.253	0.314	0.157	0.278	0.342
	100	0.257	0.452	0.551	0.288	0.477	0.557
	200	0.532	0.758	0.849	0.642	0.797	0.863
	300	0.725	0.907	0.954	0.844	0.948	0.972

Table 12: Estimated sizes of the IM tests with samples generated from the Tobit model.

Tests	Sample size	One regressor			Two regressors		
		0.01	0.05	0.10	0.01	0.05	0.10
New test	50	0.011	0.050	0.102	0.010	0.046	0.093
	100	0.010	0.047	0.094	0.010	0.048	0.097
	200	0.010	0.052	0.100	0.010	0.049	0.098
	300	0.010	0.052	0.102	0.012	0.055	0.103
IM _L	50	0.009	0.049	0.102	0.013	0.051	0.100
	100	0.010	0.053	0.098	0.010	0.053	0.100
	200	0.014	0.058	0.110	0.013	0.048	0.090
	300	0.011	0.058	0.111	0.011	0.052	0.099
IM _{DH}	50	0.011	0.052	0.099	0.011	0.050	0.098
	100	0.010	0.046	0.094	0.011	0.053	0.100
	200	0.009	0.046	0.099	0.011	0.050	0.094
	300	0.008	0.053	0.109	0.011	0.051	0.103
minp _L	50	0.009	0.049	0.098	0.009	0.049	0.099
	100	0.009	0.048	0.100	0.008	0.044	0.096
	200	0.012	0.053	0.105	0.007	0.045	0.096
	300	0.010	0.049	0.102	0.012	0.048	0.095
prod _L	50	0.009	0.049	0.093	0.008	0.048	0.096
	100	0.009	0.048	0.101	0.008	0.046	0.098
	200	0.010	0.052	0.106	0.006	0.043	0.094
	300	0.010	0.052	0.100	0.011	0.052	0.102
minp _{DH}	50	0.011	0.052	0.103	0.010	0.048	0.097
	100	0.010	0.045	0.099	0.010	0.048	0.102
	200	0.009	0.047	0.096	0.011	0.049	0.096
	300	0.011	0.054	0.103	0.012	0.050	0.103
prod _{DH}	50	0.011	0.052	0.101	0.010	0.047	0.099
	100	0.010	0.045	0.097	0.009	0.049	0.101
	200	0.010	0.046	0.098	0.011	0.049	0.100
	300	0.011	0.052	0.100	0.011	0.051	0.101

Table 13: Estimated powers of the IM tests with samples generated from (19) and (20).

Test	Sample size	Generated from (19)						Generated from (20)		
		One regressor			Two regressors			Two regressors		
		0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
New test	50	0.142	0.342	0.473	0.129	0.297	0.415	0.021	0.103	0.176
	100	0.233	0.472	0.605	0.302	0.543	0.674	0.043	0.202	0.336
	200	0.572	0.815	0.886	0.755	0.892	0.941	0.382	0.678	0.791
	300	0.806	0.939	0.972	0.933	0.984	0.994	0.717	0.915	0.960
IM _L	50	0.039	0.116	0.198	0.044	0.120	0.203	0.022	0.093	0.153
	100	0.044	0.177	0.306	0.054	0.201	0.327	0.044	0.134	0.209
	200	0.158	0.443	0.603	0.289	0.613	0.750	0.101	0.271	0.412
	300	0.341	0.659	0.803	0.653	0.888	0.943	0.193	0.456	0.605
IM _{DH}	50	0.015	0.119	0.277	0.020	0.100	0.219	0.011	0.063	0.140
	100	0.036	0.306	0.519	0.034	0.263	0.479	0.011	0.113	0.240
	200	0.292	0.718	0.855	0.279	0.787	0.905	0.064	0.369	0.577
	300	0.587	0.893	0.960	0.802	0.969	0.990	0.220	0.691	0.852
minp _L	50	0.017	0.091	0.184	0.019	0.077	0.152	0.020	0.087	0.146
	100	0.036	0.197	0.361	0.019	0.123	0.247	0.025	0.075	0.127
	200	0.187	0.605	0.776	0.178	0.538	0.717	0.023	0.103	0.196
	300	0.444	0.831	0.931	0.465	0.834	0.932	0.023	0.166	0.347
prod _L	50	0.021	0.082	0.166	0.028	0.104	0.185	0.020	0.082	0.140
	100	0.026	0.167	0.326	0.050	0.192	0.323	0.023	0.075	0.130
	200	0.131	0.494	0.683	0.354	0.674	0.807	0.033	0.124	0.233
	300	0.303	0.738	0.876	0.661	0.912	0.961	0.037	0.218	0.414
minp _{DH}	50	0.030	0.153	0.255	0.029	0.111	0.189	0.015	0.071	0.143
	100	0.083	0.325	0.479	0.047	0.192	0.344	0.015	0.089	0.184
	200	0.365	0.687	0.808	0.241	0.580	0.730	0.044	0.206	0.348
	300	0.591	0.858	0.927	0.523	0.843	0.927	0.111	0.392	0.568
prod _{DH}	50	0.040	0.130	0.202	0.030	0.105	0.172	0.015	0.059	0.118
	100	0.075	0.234	0.348	0.060	0.208	0.310	0.015	0.067	0.136
	200	0.222	0.486	0.636	0.267	0.550	0.685	0.039	0.172	0.317
	300	0.364	0.653	0.786	0.538	0.805	0.892	0.107	0.341	0.537

Online supplementary document

Our analysis in Sections 4.1.2 and 4.1.3 is based on the following full set of simulation results.

Table S1: Powers of the new tests and three competing tests estimated through parametric sampling when the regressor is X_1 and $d = 4$.

DGP	T	α	Single simulation		Double simulation		Minimum	Product	Q
			NRR	MCMC	NRR	MCMC	p	p	
AR(1): $\rho_1=0.25$	50	0.10	0.411	0.417	0.418	0.423	0.387	0.365	0.370
		0.05	0.294	0.299	0.298	0.309	0.278	0.259	0.263
		0.01	0.128	0.134	0.124	0.131	0.115	0.106	0.103
	100	0.10	0.645	0.651	0.656	0.659	0.658	0.610	0.627
		0.05	0.520	0.529	0.519	0.531	0.549	0.482	0.504
		0.01	0.276	0.288	0.268	0.282	0.326	0.259	0.277
	200	0.10	0.898	0.900	0.903	0.906	0.918	0.881	0.897
		0.05	0.835	0.838	0.838	0.847	0.868	0.805	0.833
		0.01	0.633	0.653	0.616	0.646	0.728	0.586	0.633
500	0.10	0.999	0.999	0.999	0.999	1.000	0.999	0.999	
	0.05	0.997	0.998	0.997	0.989	0.999	0.996	0.998	
	0.01	0.985	0.988	0.988	0.989	0.994	0.983	0.989	
AR(2): $\rho_1=0.05$ $\rho_2=0.10$	50	0.10	0.208	0.210	0.212	0.214	0.154	0.172	0.170
		0.05	0.125	0.126	0.125	0.126	0.088	0.105	0.104
		0.01	0.041	0.042	0.036	0.038	0.022	0.031	0.030
	100	0.10	0.256	0.261	0.263	0.266	0.214	0.235	0.234
		0.05	0.163	0.166	0.167	0.169	0.132	0.149	0.148
		0.01	0.055	0.057	0.056	0.058	0.037	0.053	0.052
	200	0.10	0.363	0.369	0.382	0.388	0.324	0.348	0.352
		0.05	0.249	0.255	0.261	0.265	0.221	0.243	0.246
		0.01	0.097	0.099	0.099	0.101	0.084	0.093	0.097
500	0.10	0.638	0.645	0.647	0.651	0.616	0.628	0.639	
	0.05	0.521	0.531	0.523	0.531	0.489	0.507	0.519	
	0.01	0.284	0.298	0.275	0.286	0.273	0.293	0.302	
AR(2): $\rho_1=0.05$ $\rho_2=0.20$	50	0.10	0.370	0.374	0.381	0.382	0.294	0.315	0.313
		0.05	0.261	0.263	0.266	0.273	0.199	0.221	0.222
		0.01	0.117	0.119	0.107	0.113	0.070	0.089	0.087
	100	0.10	0.542	0.551	0.553	0.557	0.490	0.499	0.507
		0.05	0.419	0.424	0.418	0.429	0.373	0.378	0.387
		0.01	0.214	0.222	0.196	0.210	0.176	0.189	0.197
	200	0.10	0.789	0.797	0.800	0.804	0.777	0.758	0.774
		0.05	0.692	0.701	0.698	0.704	0.683	0.650	0.672
		0.01	0.445	0.464	0.458	0.469	0.466	0.413	0.443
500	0.10	0.990	0.990	0.990	0.990	0.991	0.985	0.988	
	0.05	0.978	0.979	0.978	0.979	0.981	0.969	0.976	
	0.01	0.911	0.923	0.923	0.929	0.938	0.900	0.919	

Table S1: Powers of the new tests and three competing tests estimated through parametric sampling when the regressor is X_1 and $d = 4$ (continued).

DGP	T	α	Single simulation		Double simulation		Minimum	Product	Q	
			NRR	MCMC	NRR	MCMC	p	p		
AR(3): $\rho_1=0.10$	50	0.10	0.347	0.352	0.356	0.360	0.253	0.298	0.292	
		$\rho_2=0.10$	0.05	0.241	0.249	0.251	0.254	0.164	0.217	0.212
		$\rho_3=0.10$	0.01	0.110	0.121	0.106	0.113	0.057	0.100	0.094
	100	0.10	0.488	0.484	0.494	0.496	0.405	0.470	0.467	
		0.05	0.371	0.374	0.372	0.380	0.291	0.369	0.365	
		0.01	0.192	0.184	0.190	0.202	0.121	0.207	0.203	
	200	0.10	0.707	0.717	0.708	0.720	0.635	0.714	0.712	
		0.05	0.599	0.604	0.599	0.611	0.513	0.620	0.614	
		0.01	0.367	0.361	0.377	0.394	0.287	0.425	0.421	
	500	0.10	0.964	0.963	0.966	0.967	0.934	0.968	0.967	
		0.05	0.933	0.931	0.934	0.939	0.881	0.946	0.943	
		0.01	0.807	0.830	0.803	0.828	0.712	0.870	0.865	
AR(3): $\rho_1=0.05$	50	0.10	0.379	0.380	0.376	0.380	0.289	0.309	0.312	
		$\rho_2=0.05$	0.05	0.271	0.271	0.269	0.270	0.190	0.214	0.219
		$\rho_3=0.20$	0.01	0.122	0.121	0.117	0.120	0.064	0.086	0.083
	100	0.10	0.566	0.570	0.559	0.562	0.508	0.506	0.519	
		0.05	0.449	0.451	0.438	0.442	0.385	0.388	0.401	
		0.01	0.234	0.242	0.218	0.223	0.181	0.195	0.205	
	200	0.10	0.816	0.822	0.811	0.816	0.793	0.780	0.797	
		0.05	0.720	0.727	0.706	0.719	0.702	0.683	0.703	
		0.01	0.480	0.500	0.490	0.501	0.488	0.449	0.481	
	500	0.10	0.992	0.993	0.992	0.993	0.993	0.989	0.992	
		0.05	0.983	0.984	0.982	0.983	0.985	0.978	0.982	
		0.01	0.940	0.945	0.929	0.942	0.950	0.932	0.943	
AR(4): $\rho_1=0.05$	50	0.10	0.533	0.533	0.533	0.533	0.390	0.454	0.452	
		$\rho_2=0.10$	0.05	0.422	0.424	0.424	0.429	0.283	0.367	0.367
		$\rho_3=0.15$	0.01	0.255	0.258	0.247	0.252	0.122	0.217	0.208
	100	0.10	0.745	0.748	0.734	0.737	0.625	0.699	0.698	
		0.05	0.649	0.656	0.648	0.650	0.510	0.610	0.607	
		0.01	0.444	0.457	0.463	0.474	0.280	0.436	0.434	
	200	0.10	0.934	0.937	0.935	0.936	0.881	0.922	0.922	
		0.05	0.890	0.895	0.890	0.894	0.806	0.880	0.878	
		0.01	0.745	0.765	0.761	0.769	0.605	0.756	0.757	
	500	0.10	1.000	1.000	1.000	1.000	0.998	1.000	0.999	
		0.05	0.998	0.999	0.999	0.999	0.995	0.998	0.998	
		0.01	0.990	0.993	0.993	0.994	0.973	0.994	0.994	

Table S1: Powers of the new tests and three competing tests estimated through parametric sampling when the regressor is X_1 and $d = 4$ (continued).

DGP	T	α	Single simulation		Double simulation		Minimum	Product	Q	
			NRR	MCMC	NRR	MCMC	p	p		
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.05$ $\rho_3=0.10$	50	0.10	0.334	0.337	0.339	0.342	0.233	0.278	0.275	
		0.05	0.236	0.237	0.238	0.242	0.149	0.198	0.198	
		0.01	0.106	0.108	0.104	0.105	0.047	0.087	0.082	
	100	0.10	0.464	0.469	0.454	0.459	0.355	0.421	0.421	
			0.05	0.343	0.350	0.343	0.349	0.252	0.319	0.317
			0.01	0.164	0.172	0.165	0.167	0.096	0.171	0.169
		200	0.10	0.657	0.664	0.657	0.663	0.564	0.641	0.636
			0.05	0.542	0.551	0.548	0.556	0.438	0.539	0.535
			0.01	0.304	0.324	0.325	0.343	0.222	0.345	0.344
	500	0.10	0.936	0.939	0.936	0.939	0.890	0.935	0.935	
		0.05	0.889	0.896	0.890	0.896	0.811	0.895	0.893	
		0.01	0.731	0.750	0.747	0.756	0.603	0.783	0.776	
AR(4): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.05$ $\rho_3=0.05$	50	0.10	0.218	0.219	0.219	0.222	0.149	0.173	0.170	
		0.05	0.137	0.138	0.138	0.139	0.085	0.110	0.108	
		0.01	0.048	0.048	0.043	0.045	0.021	0.036	0.034	
	100	0.10	0.265	0.269	0.265	0.270	0.204	0.235	0.232	
			0.05	0.172	0.175	0.168	0.171	0.124	0.155	0.152
			0.01	0.061	0.064	0.060	0.063	0.033	0.062	0.060
		200	0.10	0.356	0.362	0.363	0.368	0.290	0.344	0.339
			0.05	0.249	0.253	0.246	0.252	0.189	0.247	0.243
			0.01	0.100	0.106	0.102	0.106	0.066	0.109	0.106
	500	0.10	0.611	0.613	0.602	0.611	0.513	0.612	0.606	
		0.05	0.483	0.496	0.485	0.496	0.377	0.504	0.496	
		0.01	0.250	0.264	0.273	0.298	0.170	0.314	0.302	

Table S2: Powers of the new tests and three competing tests estimated through parametric sampling when the regressor is X_1 and $d = 6$.

DGP	T	α	Single simulation		Double simulation		Minimum	Product	Q
			NRR	MCMC	NRR	MCMC	p	p	
AR(1): $\rho_1=0.25$	50	0.10	0.366	0.374	0.358	0.371	0.350	0.326	0.327
		0.05	0.256	0.269	0.249	0.265	0.247	0.225	0.228
		0.01	0.102	0.110	0.099	0.109	0.101	0.085	0.083
	100	0.10	0.566	0.582	0.568	0.591	0.616	0.539	0.560
		0.05	0.437	0.453	0.434	0.455	0.500	0.402	0.430
		0.01	0.217	0.226	0.213	0.221	0.293	0.203	0.218
	200	0.10	0.850	0.864	0.859	0.872	0.895	0.827	0.855
		0.05	0.765	0.790	0.778	0.798	0.841	0.737	0.775
		0.01	0.564	0.587	0.570	0.597	0.691	0.492	0.546
	500	0.10	0.997	0.998	0.997	0.998	0.999	0.996	0.998
		0.05	0.993	0.995	0.992	0.995	0.999	0.992	0.995
		0.01	0.974	0.977	0.968	0.973	0.994	0.961	0.977
AR(2): $\rho_1=0.05$ $\rho_2=0.10$	50	0.10	0.196	0.197	0.196	0.199	0.143	0.159	0.157
		0.05	0.117	0.121	0.121	0.124	0.082	0.093	0.091
		0.01	0.037	0.037	0.038	0.039	0.019	0.027	0.026
	100	0.10	0.223	0.233	0.237	0.244	0.195	0.210	0.209
		0.05	0.139	0.144	0.137	0.141	0.116	0.124	0.126
		0.01	0.045	0.046	0.039	0.040	0.033	0.042	0.042
	200	0.10	0.319	0.328	0.329	0.340	0.289	0.310	0.314
		0.05	0.214	0.223	0.219	0.227	0.191	0.209	0.211
		0.01	0.080	0.081	0.079	0.085	0.069	0.072	0.076
	500	0.10	0.575	0.588	0.576	0.599	0.558	0.564	0.578
		0.05	0.440	0.458	0.435	0.459	0.441	0.435	0.455
		0.01	0.225	0.234	0.217	0.232	0.240	0.222	0.241
AR(2): $\rho_1=0.05$ $\rho_2=0.20$	50	0.10	0.336	0.342	0.333	0.341	0.266	0.285	0.284
		0.05	0.233	0.240	0.227	0.238	0.177	0.193	0.193
		0.01	0.091	0.095	0.090	0.095	0.062	0.072	0.069
	100	0.10	0.472	0.490	0.475	0.486	0.447	0.442	0.449
		0.05	0.351	0.365	0.352	0.362	0.333	0.314	0.325
		0.01	0.164	0.168	0.162	0.165	0.153	0.150	0.156
	200	0.10	0.719	0.733	0.720	0.737	0.734	0.689	0.714
		0.05	0.604	0.620	0.606	0.625	0.637	0.575	0.603
		0.01	0.382	0.382	0.381	0.392	0.422	0.339	0.368
	500	0.10	0.978	0.982	0.979	0.983	0.987	0.971	0.979
		0.05	0.952	0.961	0.955	0.964	0.973	0.943	0.960
		0.01	0.859	0.870	0.864	0.876	0.925	0.835	0.874

Table S2: Powers of the new tests and three competing tests estimated through parametric sampling when the regressor is X_1 and $d = 6$ (continued).

DGP	T	α	Single simulation		Double simulation		Minimum	Product	Q	
			NRR	MCMC	NRR	MCMC	p	p		
AR(3): $\rho_1=0.10$	50	0.10	0.320	0.327	0.321	0.333	0.229	0.272	0.267	
		$\rho_2=0.10$	0.05	0.222	0.227	0.226	0.234	0.148	0.191	0.187
		$\rho_3=0.10$	0.01	0.097	0.100	0.092	0.096	0.050	0.083	0.077
	100	0.10	0.432	0.446	0.427	0.443	0.367	0.423	0.419	
		0.05	0.323	0.335	0.305	0.320	0.257	0.316	0.315	
		0.01	0.160	0.167	0.144	0.148	0.106	0.172	0.167	
	200	0.10	0.643	0.660	0.641	0.662	0.582	0.660	0.660	
		0.05	0.529	0.553	0.523	0.544	0.463	0.562	0.560	
		0.01	0.320	0.329	0.321	0.337	0.249	0.363	0.364	
	500	0.10	0.935	0.947	0.937	0.948	0.911	0.951	0.951	
		0.05	0.886	0.903	0.892	0.906	0.849	0.918	0.920	
		0.01	0.749	0.765	0.764	0.778	0.675	0.811	0.816	
AR(3): $\rho_1=0.05$	50	0.10	0.343	0.347	0.353	0.359	0.258	0.282	0.281	
		$\rho_2=0.05$	0.05	0.242	0.249	0.249	0.257	0.171	0.192	0.195
		$\rho_3=0.20$	0.01	0.102	0.105	0.106	0.108	0.057	0.075	0.072
	100	0.10	0.504	0.514	0.504	0.515	0.460	0.452	0.465	
		0.05	0.388	0.400	0.385	0.393	0.341	0.333	0.347	
		0.01	0.199	0.203	0.190	0.195	0.159	0.167	0.172	
	200	0.10	0.749	0.765	0.730	0.748	0.751	0.724	0.746	
		0.05	0.640	0.664	0.620	0.642	0.656	0.617	0.642	
		0.01	0.423	0.433	0.389	0.408	0.445	0.385	0.413	
	500	0.10	0.983	0.986	0.983	0.986	0.990	0.981	0.985	
		0.05	0.965	0.971	0.963	0.969	0.979	0.962	0.972	
		0.01	0.901	0.913	0.887	0.898	0.936	0.883	0.911	
AR(4): $\rho_1=0.05$	50	0.10	0.492	0.498	0.494	0.504	0.356	0.422	0.419	
		$\rho_2=0.10$	0.05	0.390	0.398	0.387	0.401	0.254	0.333	0.331
		$\rho_3=0.15$	0.01	0.223	0.227	0.217	0.219	0.110	0.194	0.185
	100	0.10	0.685	0.700	0.677	0.692	0.582	0.655	0.655	
		0.05	0.582	0.600	0.572	0.585	0.463	0.555	0.557	
		0.01	0.377	0.389	0.382	0.396	0.250	0.386	0.382	
	200	0.10	0.902	0.912	0.899	0.908	0.848	0.895	0.895	
		0.05	0.844	0.858	0.840	0.851	0.767	0.845	0.845	
		0.01	0.689	0.699	0.697	0.708	0.561	0.705	0.707	
	500	0.10	0.998	0.999	0.998	0.999	0.997	0.999	0.999	
		0.05	0.995	0.997	0.996	0.997	0.993	0.996	0.997	
		0.01	0.982	0.995	0.987	0.988	0.965	0.989	0.990	

Table S2: Powers of the new tests and three competing tests estimated through parametric sampling when the regressor is X_1 and $d = 6$ (continued).

DGP	T	α	Single simulation		Double simulation		Minimum	Product	Q	
			NRR	MCMC	NRR	MCMC	p	p		
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.05$ $\rho_3=0.10$	50	0.10	0.314	0.319	0.315	0.323	0.214	0.253	0.252	
		0.05	0.219	0.224	0.215	0.223	0.134	0.175	0.173	
		0.01	0.091	0.094	0.087	0.091	0.042	0.073	0.069	
	100	0.10	0.403	0.418	0.418	0.430	0.323	0.377	0.375	
			0.05	0.296	0.307	0.302	0.313	0.220	0.276	0.276
			0.01	0.138	0.142	0.139	0.146	0.082	0.143	0.141
		200	0.10	0.595	0.611	0.597	0.609	0.509	0.588	0.587
			0.05	0.473	0.497	0.488	0.505	0.392	0.487	0.485
			0.01	0.263	0.271	0.286	0.300	0.190	0.293	0.293
	500	0.10	0.899	0.912	0.898	0.908	0.854	0.909	0.909	
		0.05	0.834	0.854	0.838	0.851	0.768	0.856	0.858	
		0.01	0.666	0.685	0.683	0.694	0.561	0.714	0.718	
AR(4): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.05$ $\rho_3=0.05$	50	0.10	0.211	0.212	0.207	0.210	0.140	0.159	0.157	
		0.05	0.131	0.135	0.126	0.129	0.080	0.098	0.095	
		0.01	0.044	0.043	0.039	0.040	0.019	0.031	0.028	
	100	0.10	0.234	0.240	0.234	0.242	0.185	0.209	0.209	
			0.05	0.152	0.158	0.154	0.158	0.110	0.132	0.132
			0.01	0.051	0.052	0.050	0.051	0.029	0.051	0.050
		200	0.10	0.320	0.329	0.327	0.339	0.257	0.312	0.309
			0.05	0.217	0.230	0.222	0.233	0.164	0.221	0.217
			0.01	0.085	0.088	0.089	0.093	0.055	0.089	0.087
	500	0.10	0.541	0.561	0.550	0.565	0.452	0.559	0.556	
		0.05	0.413	0.429	0.434	0.443	0.330	0.442	0.440	
		0.01	0.213	0.221	0.240	0.245	0.144	0.254	0.251	

Table S3: Sizes and powers of the new tests and three competing tests estimated through parametric sampling when $d = 4$ and the regressors are either $X2$ or $X3$.

DGP	T	α	X2					X3				
			NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p
Size	50	0.01	0.010	0.011	0.012	0.011	0.011	0.010	0.010	0.009	0.008	0.009
		0.05	0.050	0.050	0.050	0.049	0.050	0.048	0.049	0.047	0.046	0.047
		0.10	0.096	0.097	0.096	0.094	0.099	0.092	0.092	0.095	0.098	0.096
	100	0.01	0.011	0.011	0.010	0.009	0.010	0.011	0.011	0.010	0.011	0.010
		0.05	0.051	0.051	0.050	0.049	0.051	0.048	0.049	0.051	0.048	0.051
		0.10	0.102	0.102	0.101	0.102	0.102	0.107	0.106	0.103	0.099	0.102
	200	0.01	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.009	0.009	0.009
		0.05	0.050	0.049	0.050	0.051	0.047	0.049	0.048	0.050	0.049	0.051
		0.10	0.101	0.101	0.104	0.101	0.103	0.100	0.101	0.101	0.098	0.101
500	0.01	0.008	0.008	0.010	0.010	0.010	0.010	0.009	0.010	0.011	0.010	
	0.05	0.050	0.050	0.050	0.049	0.049	0.051	0.051	0.049	0.048	0.049	
	0.10	0.100	0.100	0.101	0.101	0.102	0.099	0.099	0.101	0.100	0.101	
Power												
AR(1): $\rho_1=0.25$	50	0.01	0.071	0.075	0.038	0.017	0.038	0.059	0.063	0.037	0.015	0.039
		0.05	0.203	0.207	0.116	0.069	0.121	0.175	0.178	0.123	0.082	0.129
		0.10	0.299	0.304	0.184	0.129	0.195	0.263	0.266	0.194	0.155	0.205
	100	0.01	0.215	0.227	0.105	0.080	0.105	0.194	0.204	0.143	0.143	0.139
		0.05	0.447	0.455	0.250	0.222	0.252	0.403	0.411	0.320	0.318	0.315
		0.10	0.567	0.578	0.360	0.337	0.363	0.535	0.535	0.438	0.435	0.432
	200	0.01	0.571	0.591	0.400	0.466	0.364	0.559	0.573	0.480	0.570	0.447
		0.05	0.781	0.791	0.653	0.695	0.621	0.768	0.777	0.733	0.775	0.704
		0.10	0.863	0.872	0.766	0.790	0.741	0.853	0.859	0.825	0.850	0.806
500	0.01	0.966	0.975	0.975	0.987	0.963	0.967	0.977	0.982	0.992	0.973	
	0.05	0.996	0.997	0.994	0.997	0.991	0.996	0.997	0.996	0.998	0.994	
	0.10	0.999	0.999	0.998	0.999	0.997	0.999	0.999	0.999	0.999	0.998	
AR(2): $\rho_1=0.05$ $\rho_2=0.10$	50	0.01	0.023	0.024	0.015	0.011	0.014	0.021	0.022	0.006	0.005	0.007
		0.05	0.087	0.089	0.055	0.050	0.054	0.086	0.088	0.039	0.035	0.039
		0.10	0.153	0.154	0.100	0.091	0.101	0.149	0.150	0.077	0.074	0.079
	100	0.01	0.041	0.042	0.015	0.009	0.015	0.040	0.042	0.014	0.014	0.015
		0.05	0.131	0.131	0.057	0.048	0.058	0.131	0.134	0.065	0.055	0.067
		0.10	0.214	0.217	0.108	0.097	0.110	0.221	0.223	0.118	0.108	0.120
	200	0.01	0.076	0.079	0.028	0.022	0.027	0.081	0.084	0.038	0.033	0.039
		0.05	0.214	0.216	0.104	0.093	0.104	0.217	0.218	0.139	0.119	0.139
		0.10	0.321	0.326	0.185	0.166	0.182	0.320	0.325	0.221	0.199	0.222
500	0.01	0.261	0.276	0.186	0.167	0.177	0.275	0.287	0.223	0.199	0.214	
	0.05	0.521	0.527	0.375	0.355	0.364	0.525	0.534	0.427	0.394	0.417	
	0.10	0.635	0.640	0.503	0.483	0.491	0.639	0.643	0.551	0.522	0.541	

Table S3: Sizes and powers of the new tests and three competing tests estimated through parametric sampling when $d = 4$ and the regressors are either $X2$ or $X3$ (continued).

DGP	T	α	$X2$					$X3$				
			NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p
AR(2): $\rho_1=0.05$ $\rho_2=0.20$	50	0.01	0.053	0.054	0.035	0.021	0.036	0.055	0.057	0.008	0.006	0.007
		0.05	0.165	0.166	0.094	0.071	0.096	0.166	0.168	0.045	0.041	0.044
		0.10	0.262	0.265	0.150	0.124	0.156	0.264	0.267	0.087	0.085	0.088
	100	0.01	0.151	0.156	0.049	0.031	0.048	0.156	0.163	0.052	0.049	0.050
		0.05	0.335	0.336	0.144	0.115	0.144	0.340	0.349	0.158	0.146	0.158
		0.10	0.448	0.452	0.225	0.198	0.230	0.470	0.474	0.252	0.235	0.249
	200	0.01	0.357	0.377	0.196	0.208	0.179	0.382	0.398	0.244	0.259	0.230
		0.05	0.617	0.623	0.407	0.421	0.387	0.627	0.634	0.474	0.481	0.457
		0.10	0.735	0.740	0.541	0.540	0.524	0.744	0.749	0.597	0.601	0.583
500	0.01	0.884	0.903	0.841	0.875	0.812	0.892	0.907	0.869	0.896	0.840	
	0.05	0.974	0.977	0.946	0.955	0.931	0.974	0.978	0.956	0.962	0.944	
	0.10	0.988	0.989	0.971	0.977	0.965	0.989	0.990	0.978	0.981	0.972	
AR(3): $\rho_1=0.10$ $\rho_2=0.10$ $\rho_3=0.10$	50	0.01	0.030	0.030	0.008	0.007	0.008	0.032	0.033	0.006	0.005	0.006
		0.05	0.104	0.105	0.036	0.033	0.036	0.111	0.112	0.036	0.033	0.035
		0.10	0.176	0.178	0.074	0.070	0.076	0.185	0.186	0.073	0.071	0.075
	100	0.01	0.095	0.097	0.014	0.009	0.015	0.096	0.100	0.036	0.023	0.038
		0.05	0.227	0.228	0.056	0.048	0.059	0.231	0.237	0.109	0.088	0.114
		0.10	0.324	0.327	0.105	0.092	0.108	0.339	0.340	0.182	0.155	0.186
	200	0.01	0.255	0.270	0.123	0.069	0.127	0.272	0.284	0.202	0.116	0.208
		0.05	0.474	0.479	0.268	0.199	0.271	0.475	0.484	0.383	0.291	0.392
		0.10	0.593	0.600	0.366	0.302	0.371	0.598	0.606	0.488	0.414	0.498
500	0.01	0.766	0.780	0.716	0.520	0.723	0.773	0.790	0.783	0.604	0.787	
	0.05	0.917	0.919	0.855	0.753	0.860	0.919	0.923	0.897	0.812	0.901	
	0.10	0.948	0.949	0.906	0.847	0.909	0.952	0.953	0.937	0.888	0.939	
AR(3): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.20$	50	0.01	0.037	0.037	0.026	0.011	0.027	0.032	0.034	0.021	0.009	0.022
		0.05	0.128	0.127	0.081	0.051	0.089	0.121	0.122	0.081	0.050	0.084
		0.10	0.209	0.210	0.139	0.100	0.150	0.198	0.198	0.138	0.105	0.144
	100	0.01	0.124	0.124	0.036	0.023	0.036	0.114	0.119	0.055	0.048	0.055
		0.05	0.294	0.292	0.117	0.094	0.117	0.271	0.277	0.168	0.150	0.166
		0.10	0.412	0.411	0.197	0.171	0.201	0.395	0.397	0.262	0.243	0.258
	200	0.01	0.373	0.385	0.182	0.199	0.164	0.372	0.382	0.256	0.280	0.237
		0.05	0.608	0.615	0.394	0.408	0.371	0.596	0.605	0.505	0.509	0.482
		0.10	0.725	0.729	0.526	0.530	0.505	0.720	0.724	0.628	0.634	0.611
500	0.01	0.904	0.911	0.864	0.881	0.836	0.903	0.909	0.903	0.916	0.879	
	0.05	0.974	0.975	0.953	0.959	0.942	0.973	0.974	0.968	0.973	0.960	
	0.10	0.988	0.988	0.976	0.979	0.970	0.987	0.988	0.984	0.987	0.980	

Table S3: Sizes and powers of the new tests and three competing tests estimated through parametric sampling when $d = 4$ and the regressors are either $X2$ or $X3$ (continued).

DGP	T	α	$X2$					$X3$					
			NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p	
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.15$ $\rho_4=0.15$	50	0.01	0.033	0.033	0.017	0.013	0.017	0.062	0.063	0.008	0.006	0.008	
		0.05	0.118	0.117	0.060	0.055	0.064	0.172	0.173	0.042	0.035	0.042	
		0.10	0.196	0.197	0.108	0.097	0.115	0.262	0.261	0.078	0.075	0.081	
	100	0.01	0.156	0.156	0.025	0.013	0.027	0.215	0.224	0.070	0.041	0.073	
		0.05	0.332	0.329	0.083	0.062	0.086	0.408	0.416	0.179	0.133	0.183	
		0.10	0.448	0.447	0.142	0.122	0.147	0.537	0.537	0.266	0.221	0.270	
		200	0.01	0.524	0.539	0.290	0.168	0.292	0.593	0.607	0.441	0.285	0.450
			0.05	0.733	0.738	0.483	0.380	0.485	0.785	0.790	0.653	0.534	0.656
			0.10	0.831	0.832	0.594	0.515	0.595	0.860	0.863	0.745	0.664	0.751
	500	0.01	0.977	0.980	0.966	0.892	0.966	0.983	0.985	0.981	0.934	0.981	
		0.05	0.995	0.996	0.989	0.970	0.989	0.997	0.997	0.994	0.983	0.994	
		0.10	0.998	0.998	0.994	0.986	0.994	0.999	0.999	0.997	0.993	0.997	
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.05$ $\rho_4=0.10$	50	0.01	0.027	0.027	0.018	0.012	0.018	0.033	0.033	0.004	0.004	0.004	
		0.05	0.098	0.097	0.056	0.049	0.058	0.112	0.113	0.027	0.025	0.027	
		0.10	0.165	0.166	0.097	0.088	0.103	0.183	0.183	0.054	0.056	0.056	
	100	0.01	0.074	0.075	0.019	0.009	0.020	0.089	0.094	0.023	0.015	0.023	
		0.05	0.195	0.194	0.056	0.043	0.058	0.221	0.227	0.074	0.061	0.076	
		0.10	0.289	0.290	0.098	0.087	0.102	0.329	0.328	0.129	0.114	0.131	
		200	0.01	0.198	0.206	0.080	0.044	0.081	0.226	0.235	0.130	0.074	0.135
			0.05	0.396	0.400	0.192	0.149	0.194	0.428	0.430	0.281	0.208	0.287
			0.10	0.519	0.523	0.280	0.233	0.282	0.548	0.550	0.378	0.316	0.384
	500	0.01	0.638	0.659	0.575	0.397	0.578	0.672	0.691	0.646	0.462	0.652	
		0.05	0.858	0.861	0.748	0.643	0.751	0.865	0.871	0.804	0.697	0.812	
		0.10	0.911	0.913	0.830	0.757	0.831	0.920	0.922	0.872	0.802	0.874	
AR(4): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.05$ $\rho_4=0.05$	50	0.01	0.017	0.017	0.010	0.009	0.010	0.017	0.018	0.005	0.005	0.005	
		0.05	0.071	0.069	0.041	0.039	0.042	0.071	0.072	0.033	0.032	0.033	
		0.10	0.126	0.127	0.078	0.078	0.082	0.128	0.128	0.066	0.069	0.068	
	100	0.01	0.031	0.031	0.007	0.005	0.008	0.033	0.035	0.010	0.008	0.010	
		0.05	0.106	0.104	0.032	0.028	0.034	0.108	0.111	0.042	0.039	0.043	
		0.10	0.178	0.178	0.065	0.065	0.068	0.192	0.192	0.086	0.080	0.087	
		200	0.01	0.060	0.063	0.017	0.010	0.017	0.067	0.069	0.031	0.019	0.034
			0.05	0.176	0.178	0.061	0.052	0.061	0.184	0.187	0.105	0.083	0.108
			0.10	0.274	0.277	0.115	0.102	0.115	0.281	0.283	0.174	0.152	0.179
	500	0.01	0.200	0.209	0.146	0.076	0.150	0.215	0.226	0.199	0.111	0.203	
		0.05	0.441	0.445	0.295	0.219	0.302	0.450	0.456	0.368	0.274	0.376	
		0.10	0.557	0.561	0.401	0.335	0.405	0.566	0.571	0.482	0.404	0.484	

Table S4: Sizes and powers of the new tests and three competing tests estimated via bootstrap sampling when $d = 4$, the error term is either normal or mixture normal, and the regressor is $X1$.

DGP	T	α	Normal errors					Mixture normal errors				
			NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p
Size	50	0.01	0.007	0.008	0.009	0.008	0.009	0.010	0.010	0.011	0.013	0.012
		0.05	0.044	0.043	0.044	0.046	0.044	0.049	0.048	0.047	0.051	0.047
		0.10	0.090	0.090	0.093	0.090	0.091	0.095	0.094	0.096	0.099	0.094
	100	0.01	0.010	0.010	0.011	0.011	0.011	0.010	0.010	0.010	0.009	0.010
		0.05	0.046	0.046	0.047	0.046	0.049	0.047	0.048	0.048	0.049	0.049
		0.10	0.097	0.098	0.097	0.098	0.097	0.099	0.098	0.100	0.099	0.102
	200	0.01	0.011	0.011	0.011	0.011	0.011	0.012	0.012	0.010	0.011	0.011
		0.05	0.050	0.049	0.049	0.051	0.048	0.053	0.053	0.051	0.052	0.051
		0.10	0.101	0.101	0.098	0.100	0.101	0.104	0.104	0.099	0.097	0.098
500	0.01	0.010	0.010	0.011	0.011	0.011	0.009	0.009	0.009	0.009	0.009	
	0.05	0.049	0.049	0.050	0.049	0.050	0.052	0.051	0.046	0.047	0.046	
	0.10	0.099	0.100	0.099	0.099	0.099	0.100	0.099	0.092	0.093	0.094	
Power												
AR(1): $\rho_1=0.25$	50	0.01	0.057	0.064	0.067	0.077	0.067	0.076	0.082	0.075	0.087	0.076
		0.05	0.191	0.198	0.194	0.209	0.189	0.214	0.222	0.215	0.219	0.214
		0.10	0.298	0.304	0.300	0.309	0.294	0.329	0.337	0.314	0.323	0.310
	100	0.01	0.215	0.233	0.231	0.288	0.219	0.122	0.125	0.069	0.176	0.060
		0.05	0.460	0.472	0.439	0.481	0.422	0.302	0.304	0.252	0.365	0.228
		0.10	0.586	0.596	0.569	0.600	0.553	0.429	0.430	0.396	0.479	0.373
	200	0.01	0.565	0.579	0.600	0.679	0.563	0.595	0.611	0.589	0.676	0.543
		0.05	0.798	0.804	0.800	0.846	0.772	0.807	0.818	0.801	0.848	0.772
		0.10	0.877	0.876	0.877	0.901	0.858	0.891	0.895	0.876	0.903	0.854
500	0.01	0.987	0.988	0.989	0.994	0.983	0.981	0.984	0.989	0.995	0.982	
	0.05	0.997	0.998	0.997	0.999	0.996	0.998	0.998	0.998	0.999	0.996	
	0.10	0.999	0.999	0.999	1.000	0.999	0.999	0.999	0.999	1.000	0.999	
AR(2): $\rho_1=0.05$ $\rho_2=0.10$	50	0.01	0.017	0.019	0.019	0.014	0.019	0.018	0.020	0.018	0.016	0.019
		0.05	0.077	0.078	0.071	0.065	0.070	0.082	0.083	0.078	0.071	0.079
		0.10	0.137	0.138	0.134	0.120	0.133	0.148	0.150	0.141	0.129	0.142
	100	0.01	0.034	0.037	0.038	0.030	0.039	0.028	0.030	0.034	0.029	0.035
		0.05	0.129	0.132	0.115	0.102	0.118	0.114	0.116	0.114	0.106	0.115
		0.10	0.213	0.217	0.195	0.180	0.195	0.200	0.202	0.195	0.187	0.196
	200	0.01	0.072	0.074	0.081	0.063	0.079	0.067	0.069	0.078	0.064	0.073
		0.05	0.197	0.198	0.204	0.193	0.202	0.211	0.217	0.210	0.192	0.207
		0.10	0.303	0.304	0.313	0.291	0.310	0.319	0.324	0.309	0.291	0.306
500	0.01	0.270	0.275	0.298	0.261	0.287	0.256	0.257	0.286	0.255	0.273	
	0.05	0.486	0.496	0.507	0.473	0.495	0.488	0.495	0.497	0.470	0.488	
	0.10	0.618	0.622	0.619	0.595	0.610	0.627	0.631	0.622	0.592	0.612	

Table S4: Sizes and powers of the new tests and three competing tests estimated via bootstrap sampling when $d = 4$, the error term is either normal or mixture normal, and the regressor is $X1$ (continued).

DGP	T	α	Normal errors					Mixture normal errors				
			NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p
Power												
AR(2): $\rho_1=0.05$ $\rho_2=0.20$	50	0.01	0.051	0.055	0.052	0.043	0.053	0.054	0.056	0.057	0.055	0.057
		0.05	0.166	0.170	0.152	0.143	0.150	0.161	0.164	0.161	0.148	0.161
		0.10	0.257	0.262	0.244	0.227	0.240	0.259	0.262	0.248	0.238	0.244
	100	0.01	0.145	0.155	0.151	0.147	0.148	0.140	0.147	0.149	0.147	0.148
		0.05	0.340	0.349	0.317	0.308	0.311	0.321	0.327	0.331	0.322	0.322
		0.10	0.469	0.477	0.437	0.423	0.430	0.444	0.447	0.450	0.437	0.439
	200	0.01	0.380	0.390	0.397	0.405	0.379	0.377	0.392	0.400	0.407	0.372
		0.05	0.611	0.617	0.619	0.642	0.597	0.653	0.657	0.628	0.646	0.605
		0.10	0.732	0.734	0.734	0.743	0.717	0.762	0.760	0.735	0.747	0.718
500	0.01	0.904	0.911	0.916	0.932	0.896	0.848	0.859	0.893	0.920	0.868	
	0.05	0.972	0.973	0.973	0.978	0.966	0.957	0.958	0.967	0.975	0.959	
	0.10	0.987	0.987	0.986	0.989	0.983	0.979	0.979	0.984	0.988	0.980	
AR(3): $\rho_1=0.10$ $\rho_2=0.10$ $\rho_3=0.10$	50	0.01	0.037	0.039	0.046	0.029	0.048	0.016	0.016	0.025	0.025	0.025
		0.05	0.122	0.126	0.123	0.103	0.124	0.069	0.070	0.095	0.089	0.096
		0.10	0.199	0.202	0.196	0.174	0.200	0.126	0.125	0.161	0.154	0.163
	100	0.01	0.124	0.133	0.143	0.091	0.148	0.124	0.129	0.139	0.091	0.140
		0.05	0.293	0.299	0.274	0.220	0.279	0.282	0.287	0.277	0.228	0.279
		0.10	0.397	0.401	0.372	0.324	0.377	0.388	0.395	0.373	0.328	0.379
	200	0.01	0.295	0.306	0.364	0.225	0.373	0.307	0.322	0.355	0.235	0.365
		0.05	0.501	0.506	0.545	0.455	0.551	0.516	0.525	0.549	0.447	0.556
		0.10	0.623	0.627	0.650	0.577	0.655	0.628	0.635	0.646	0.573	0.650
500	0.01	0.802	0.815	0.854	0.690	0.859	0.813	0.823	0.844	0.690	0.849	
	0.05	0.922	0.928	0.935	0.867	0.938	0.922	0.925	0.934	0.868	0.937	
	0.10	0.958	0.959	0.960	0.924	0.962	0.956	0.958	0.962	0.927	0.964	
AR(3): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.20$	50	0.01	0.047	0.048	0.042	0.035	0.042	0.058	0.060	0.053	0.048	0.053
		0.05	0.147	0.149	0.135	0.126	0.132	0.172	0.171	0.147	0.141	0.144
		0.10	0.234	0.236	0.224	0.205	0.219	0.258	0.259	0.232	0.218	0.225
	100	0.01	0.162	0.169	0.147	0.144	0.142	0.135	0.136	0.143	0.131	0.135
		0.05	0.362	0.367	0.312	0.303	0.303	0.314	0.318	0.309	0.306	0.299
		0.10	0.483	0.484	0.435	0.427	0.423	0.452	0.451	0.434	0.430	0.424
	200	0.01	0.378	0.397	0.423	0.420	0.402	0.354	0.371	0.410	0.428	0.396
		0.05	0.624	0.632	0.644	0.658	0.623	0.603	0.613	0.635	0.642	0.614
		0.10	0.743	0.748	0.750	0.754	0.735	0.721	0.724	0.742	0.738	0.725
500	0.01	0.919	0.932	0.939	0.942	0.925	0.874	0.888	0.928	0.934	0.912	
	0.05	0.981	0.981	0.980	0.983	0.976	0.967	0.971	0.978	0.980	0.973	
	0.10	0.990	0.991	0.990	0.992	0.987	0.985	0.985	0.988	0.990	0.986	

Table S4: Sizes and powers of the new tests and three competing tests estimated via bootstrap sampling when $d = 4$, the error term is either normal or mixture normal, and the regressor is $X1$ (continued).

DGP	T	α	Normal errors					Mixture normal errors					
			NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p	
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.15$ $\rho_4=0.15$	50	0.01	0.081	0.085	0.097	0.052	0.100	0.116	0.117	0.108	0.061	0.112	
		0.05	0.219	0.222	0.211	0.163	0.210	0.261	0.265	0.222	0.171	0.223	
		0.10	0.316	0.318	0.299	0.251	0.299	0.373	0.377	0.310	0.268	0.312	
	100	0.01	0.306	0.319	0.311	0.201	0.318	0.294	0.302	0.303	0.200	0.306	
		0.05	0.525	0.531	0.479	0.387	0.485	0.506	0.516	0.492	0.398	0.495	
		0.10	0.632	0.636	0.585	0.512	0.589	0.626	0.633	0.596	0.519	0.595	
		200	0.01	0.625	0.645	0.691	0.509	0.696	0.635	0.656	0.694	0.523	0.697
			0.05	0.797	0.803	0.825	0.748	0.827	0.816	0.819	0.827	0.743	0.830
			0.10	0.873	0.877	0.887	0.835	0.886	0.884	0.885	0.885	0.832	0.885
	500	0.01	0.987	0.989	0.992	0.967	0.992	0.962	0.968	0.986	0.953	0.986	
		0.05	0.997	0.997	0.998	0.993	0.998	0.990	0.991	0.996	0.989	0.997	
		0.10	0.999	0.999	0.999	0.997	0.999	0.996	0.996	0.999	0.996	0.999	
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.05$ $\rho_4=0.10$	50	0.01	0.034	0.035	0.040	0.024	0.042	0.049	0.050	0.048	0.029	0.050	
		0.05	0.117	0.119	0.115	0.093	0.115	0.148	0.149	0.123	0.100	0.124	
		0.10	0.194	0.194	0.183	0.159	0.184	0.238	0.241	0.194	0.172	0.196	
	100	0.01	0.104	0.111	0.115	0.070	0.118	0.121	0.125	0.123	0.072	0.125	
		0.05	0.247	0.254	0.229	0.182	0.235	0.263	0.266	0.238	0.192	0.240	
		0.10	0.354	0.360	0.325	0.279	0.328	0.362	0.368	0.328	0.287	0.332	
		200	0.01	0.219	0.229	0.284	0.166	0.293	0.260	0.272	0.290	0.173	0.292
			0.05	0.411	0.415	0.461	0.379	0.466	0.483	0.488	0.468	0.383	0.471
			0.10	0.541	0.543	0.568	0.501	0.572	0.603	0.604	0.575	0.504	0.577
	500	0.01	0.709	0.718	0.759	0.577	0.765	0.691	0.715	0.747	0.570	0.755	
		0.05	0.855	0.864	0.877	0.789	0.881	0.866	0.872	0.872	0.787	0.875	
		0.10	0.914	0.918	0.922	0.872	0.923	0.924	0.927	0.921	0.871	0.924	
AR(4): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.05$ $\rho_4=0.05$	50	0.01	0.015	0.016	0.016	0.011	0.017	0.020	0.021	0.020	0.017	0.020	
		0.05	0.068	0.069	0.063	0.057	0.063	0.079	0.079	0.071	0.065	0.072	
		0.10	0.124	0.125	0.120	0.107	0.117	0.140	0.142	0.124	0.119	0.124	
	100	0.01	0.035	0.038	0.040	0.024	0.041	0.029	0.030	0.035	0.023	0.037	
		0.05	0.121	0.123	0.107	0.089	0.109	0.100	0.104	0.102	0.085	0.104	
		0.10	0.200	0.203	0.177	0.159	0.179	0.177	0.178	0.170	0.155	0.171	
		200	0.01	0.061	0.063	0.081	0.046	0.085	0.049	0.052	0.081	0.052	0.086
			0.05	0.167	0.168	0.194	0.155	0.198	0.148	0.150	0.194	0.153	0.197
			0.10	0.267	0.268	0.287	0.247	0.290	0.239	0.241	0.281	0.243	0.283
	500	0.01	0.238	0.245	0.285	0.156	0.295	0.173	0.183	0.270	0.151	0.280	
		0.05	0.437	0.447	0.470	0.354	0.479	0.373	0.380	0.459	0.353	0.468	
		0.10	0.559	0.566	0.574	0.483	0.580	0.489	0.493	0.565	0.481	0.574	

Table S5: Sizes and powers of the new tests and three competing tests estimated via bootstrap sampling when $d = 4$, the error term is either normal or mixture normal, and the regressors are X_2 .

DGP	T	α	Normal errors					Mixture normal errors					
			NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p	
Size	50	0.01	0.009	0.009	0.007	0.009	0.007	0.009	0.010	0.009	0.010	0.009	
		0.05	0.047	0.047	0.039	0.048	0.041	0.048	0.048	0.048	0.046	0.048	0.046
		0.10	0.101	0.101	0.086	0.092	0.082	0.095	0.094	0.095	0.095	0.095	0.095
	100	0.01	0.007	0.007	0.009	0.008	0.008	0.020	0.021	0.020	0.015	0.020	
		0.05	0.048	0.051	0.046	0.047	0.046	0.081	0.082	0.068	0.063	0.070	
		0.10	0.116	0.114	0.095	0.105	0.095	0.145	0.147	0.127	0.120	0.127	
	200	0.01	0.009	0.009	0.013	0.014	0.011	0.011	0.011	0.011	0.009	0.011	
		0.05	0.051	0.051	0.054	0.057	0.054	0.052	0.050	0.050	0.050	0.050	
		0.10	0.099	0.100	0.106	0.102	0.105	0.103	0.102	0.101	0.101	0.100	
	500	0.01	0.008	0.009	0.008	0.005	0.008	0.011	0.010	0.010	0.009	0.011	
		0.05	0.054	0.054	0.048	0.047	0.048	0.048	0.048	0.051	0.050	0.050	
		0.10	0.102	0.104	0.099	0.101	0.098	0.097	0.098	0.099	0.098	0.100	
Power													
AR(1): $\rho_1=0.25$	50	0.01	0.146	0.148	0.024	0.009	0.025	0.151	0.157	0.024	0.011	0.026	
		0.05	0.331	0.336	0.083	0.048	0.088	0.341	0.345	0.086	0.051	0.092	
		0.10	0.469	0.476	0.136	0.093	0.145	0.473	0.477	0.143	0.099	0.151	
	100	0.01	0.291	0.302	0.096	0.104	0.094	0.284	0.300	0.087	0.090	0.085	
		0.05	0.541	0.551	0.249	0.248	0.246	0.541	0.548	0.239	0.232	0.233	
		0.10	0.677	0.685	0.357	0.362	0.349	0.673	0.679	0.350	0.350	0.345	
	200	0.01	0.603	0.629	0.468	0.539	0.425	0.626	0.644	0.445	0.516	0.407	
		0.05	0.835	0.848	0.698	0.750	0.666	0.822	0.828	0.682	0.731	0.643	
		0.10	0.909	0.910	0.801	0.831	0.775	0.892	0.896	0.784	0.822	0.760	
	500	0.01	0.974	0.983	0.980	0.990	0.968	0.982	0.985	0.978	0.990	0.969	
		0.05	0.998	0.998	0.996	0.998	0.993	0.997	0.998	0.996	0.998	0.993	
		0.10	0.999	0.999	0.998	0.999	0.998	0.999	0.999	0.998	0.999	0.997	
AR(2): $\rho_1=0.05$ $\rho_2=0.10$	50	0.01	0.075	0.075	0.004	0.003	0.003	0.068	0.068	0.004	0.004	0.004	
		0.05	0.188	0.189	0.023	0.019	0.022	0.186	0.186	0.024	0.022	0.024	
		0.10	0.292	0.294	0.046	0.041	0.045	0.288	0.290	0.049	0.046	0.049	
	100	0.01	0.069	0.069	0.008	0.007	0.009	0.057	0.059	0.007	0.006	0.007	
		0.05	0.197	0.200	0.041	0.033	0.041	0.180	0.182	0.036	0.031	0.037	
		0.10	0.303	0.305	0.079	0.073	0.080	0.286	0.290	0.073	0.068	0.075	
	200	0.01	0.090	0.095	0.035	0.026	0.035	0.095	0.099	0.032	0.025	0.032	
		0.05	0.258	0.266	0.117	0.102	0.117	0.233	0.236	0.105	0.094	0.104	
		0.10	0.374	0.376	0.196	0.176	0.193	0.341	0.347	0.175	0.165	0.174	
	500	0.01	0.296	0.304	0.205	0.173	0.194	0.280	0.284	0.200	0.171	0.192	
		0.05	0.523	0.532	0.409	0.375	0.398	0.524	0.530	0.409	0.376	0.400	
		0.10	0.638	0.648	0.532	0.499	0.523	0.649	0.652	0.529	0.500	0.523	

Table S5: Sizes and powers of the new tests and three competing tests estimated via bootstrap sampling when $d = 4$, the error term is either normal or mixture normal, and the regressors are X_2 (continued).

DGP	T	α	Normal errors					Mixture normal errors				
			NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p
AR(2): $\rho_1=0.05$ $\rho_2=0.20$	50	0.01	0.158	0.159	0.005	0.004	0.004	0.147	0.149	0.004	0.004	0.004
		0.05	0.312	0.312	0.025	0.023	0.024	0.306	0.307	0.025	0.024	0.025
		0.10	0.432	0.434	0.052	0.049	0.053	0.420	0.422	0.052	0.050	0.052
	100	0.01	0.217	0.219	0.031	0.032	0.030	0.180	0.183	0.027	0.024	0.026
		0.05	0.424	0.427	0.112	0.103	0.112	0.402	0.409	0.100	0.089	0.098
		0.10	0.546	0.550	0.185	0.179	0.184	0.537	0.543	0.171	0.164	0.171
	200	0.01	0.385	0.407	0.236	0.235	0.214	0.421	0.434	0.210	0.215	0.194
		0.05	0.661	0.671	0.435	0.447	0.419	0.643	0.648	0.417	0.429	0.394
		0.10	0.772	0.777	0.563	0.570	0.544	0.750	0.754	0.543	0.551	0.523
	500	0.01	0.911	0.918	0.856	0.876	0.823	0.903	0.909	0.854	0.879	0.830
		0.05	0.973	0.976	0.952	0.959	0.940	0.976	0.976	0.953	0.959	0.941
		0.10	0.988	0.989	0.975	0.978	0.969	0.988	0.988	0.974	0.979	0.969
AR(3): $\rho_1=0.10$ $\rho_2=0.10$ $\rho_3=0.10$	50	0.01	0.124	0.123	0.003	0.003	0.003	0.115	0.117	0.003	0.004	0.004
		0.05	0.273	0.274	0.020	0.018	0.020	0.270	0.270	0.023	0.020	0.023
		0.10	0.393	0.394	0.042	0.039	0.044	0.383	0.385	0.045	0.042	0.046
	100	0.01	0.184	0.186	0.022	0.015	0.023	0.144	0.148	0.021	0.012	0.023
		0.05	0.374	0.378	0.076	0.057	0.079	0.345	0.350	0.070	0.052	0.073
		0.10	0.498	0.502	0.130	0.110	0.135	0.478	0.480	0.123	0.103	0.127
	200	0.01	0.350	0.366	0.194	0.102	0.195	0.334	0.348	0.175	0.091	0.181
		0.05	0.583	0.593	0.351	0.263	0.359	0.562	0.569	0.332	0.244	0.336
		0.10	0.697	0.704	0.456	0.379	0.462	0.676	0.680	0.437	0.360	0.442
	500	0.01	0.827	0.837	0.767	0.561	0.767	0.807	0.815	0.757	0.567	0.764
		0.05	0.930	0.935	0.891	0.797	0.895	0.927	0.933	0.885	0.793	0.889
		0.10	0.961	0.964	0.932	0.877	0.935	0.962	0.964	0.929	0.875	0.931
AR(3): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.20$	50	0.01	0.126	0.123	0.014	0.005	0.014	0.122	0.123	0.015	0.007	0.017
		0.05	0.294	0.293	0.050	0.029	0.054	0.289	0.287	0.058	0.034	0.063
		0.10	0.427	0.426	0.091	0.060	0.096	0.419	0.419	0.101	0.068	0.106
	100	0.01	0.211	0.214	0.031	0.031	0.032	0.173	0.172	0.031	0.025	0.031
		0.05	0.446	0.448	0.117	0.104	0.116	0.402	0.404	0.115	0.095	0.113
		0.10	0.569	0.571	0.196	0.184	0.194	0.544	0.543	0.192	0.174	0.189
	200	0.01	0.436	0.456	0.245	0.254	0.220	0.462	0.474	0.226	0.236	0.207
		0.05	0.698	0.705	0.462	0.476	0.442	0.695	0.699	0.446	0.457	0.417
		0.10	0.800	0.803	0.594	0.602	0.569	0.786	0.789	0.571	0.582	0.547
	500	0.01	0.934	0.940	0.892	0.899	0.863	0.919	0.929	0.885	0.893	0.862
		0.05	0.981	0.982	0.965	0.970	0.957	0.982	0.984	0.963	0.968	0.954
		0.10	0.992	0.992	0.982	0.985	0.978	0.991	0.992	0.981	0.984	0.976

Table S5: Sizes and powers of the new tests and three competing tests estimated via bootstrap sampling when $d = 4$, the error term is either normal or mixture normal, and the regressors are X_2 (continued).

AR(4):	50	0.01	0.213	0.208	0.005	0.004	0.005	0.198	0.195	0.006	0.006	0.006
$\rho_1=0.05$		0.05	0.389	0.384	0.025	0.019	0.025	0.369	0.365	0.027	0.025	0.029
$\rho_2=0.10$		0.10	0.502	0.500	0.049	0.042	0.049	0.490	0.487	0.052	0.049	0.052
$\rho_3=0.15$	100	0.01	0.321	0.323	0.047	0.026	0.049	0.288	0.290	0.046	0.023	0.047
$\rho_4=0.15$		0.05	0.549	0.551	0.133	0.090	0.137	0.506	0.509	0.121	0.083	0.125
		0.10	0.659	0.659	0.205	0.164	0.207	0.632	0.635	0.188	0.153	0.193
	200	0.01	0.658	0.676	0.431	0.255	0.433	0.644	0.661	0.411	0.240	0.417
		0.05	0.845	0.848	0.620	0.499	0.623	0.825	0.829	0.597	0.477	0.598
		0.10	0.900	0.901	0.720	0.631	0.721	0.889	0.889	0.695	0.607	0.698
	500	0.01	0.989	0.991	0.979	0.917	0.978	0.987	0.988	0.975	0.914	0.976
		0.05	0.997	0.998	0.993	0.981	0.994	0.997	0.998	0.993	0.979	0.994
		0.10	0.999	0.999	0.997	0.992	0.997	0.999	0.999	0.997	0.991	0.997
AR(4):	50	0.01	0.127	0.126	0.002	0.002	0.002	0.114	0.113	0.003	0.003	0.003
$\rho_1=0.05$		0.05	0.268	0.266	0.014	0.012	0.015	0.256	0.253	0.017	0.017	0.017
$\rho_2=0.10$		0.10	0.374	0.373	0.032	0.030	0.033	0.359	0.359	0.034	0.034	0.034
$\rho_3=0.05$	100	0.01	0.142	0.144	0.013	0.008	0.014	0.117	0.118	0.013	0.007	0.014
$\rho_4=0.10$		0.05	0.313	0.317	0.050	0.038	0.053	0.282	0.287	0.046	0.034	0.047
		0.10	0.429	0.431	0.090	0.079	0.091	0.408	0.411	0.081	0.071	0.083
	200	0.01	0.266	0.279	0.124	0.063	0.127	0.259	0.269	0.113	0.060	0.115
		0.05	0.501	0.508	0.251	0.185	0.259	0.466	0.471	0.237	0.173	0.238
		0.10	0.617	0.620	0.350	0.286	0.352	0.584	0.589	0.324	0.269	0.328
	500	0.01	0.725	0.740	0.627	0.422	0.629	0.697	0.707	0.620	0.418	0.631
		0.05	0.878	0.885	0.793	0.679	0.801	0.868	0.876	0.787	0.675	0.792
		0.10	0.925	0.928	0.863	0.784	0.866	0.923	0.927	0.856	0.782	0.859
AR(4):	50	0.01	0.075	0.075	0.003	0.003	0.003	0.066	0.066	0.004	0.004	0.004
$\rho_1=0.05$		0.05	0.189	0.189	0.019	0.017	0.019	0.182	0.182	0.021	0.021	0.021
$\rho_2=0.05$		0.10	0.290	0.290	0.038	0.037	0.039	0.284	0.283	0.042	0.042	0.043
$\rho_3=0.05$	100	0.01	0.067	0.067	0.005	0.004	0.005	0.054	0.054	0.005	0.004	0.005
$\rho_4=0.05$		0.05	0.191	0.193	0.025	0.022	0.025	0.169	0.172	0.024	0.022	0.026
		0.10	0.291	0.292	0.053	0.052	0.055	0.271	0.273	0.053	0.050	0.054
	200	0.01	0.093	0.097	0.030	0.016	0.031	0.089	0.093	0.026	0.015	0.028
		0.05	0.248	0.253	0.087	0.071	0.091	0.228	0.230	0.082	0.065	0.083
		0.10	0.361	0.363	0.152	0.132	0.153	0.333	0.337	0.137	0.121	0.139
	500	0.01	0.274	0.282	0.183	0.091	0.186	0.248	0.255	0.179	0.091	0.188
		0.05	0.483	0.493	0.354	0.257	0.362	0.476	0.488	0.354	0.255	0.360
		0.10	0.601	0.609	0.464	0.379	0.471	0.603	0.609	0.459	0.380	0.463

Table S6: Sizes and powers of the new tests and three competing tests estimated via bootstrap sampling when $d = 4$, the error term is either normal or mixture normal, and the regressors are X_3 .

DGP	T	α	Normal errors					Mixture normal errors					
			NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p	
Size	50	0.01	0.010	0.010	0.011	0.010	0.011	0.009	0.009	0.009	0.011	0.009	
		0.05	0.049	0.049	0.052	0.051	0.052	0.048	0.048	0.047	0.047	0.047	
		0.10	0.105	0.104	0.101	0.105	0.104	0.095	0.095	0.097	0.094	0.097	
	100	0.01	0.008	0.008	0.009	0.010	0.009	0.010	0.010	0.010	0.011	0.012	0.011
		0.05	0.050	0.050	0.048	0.047	0.048	0.049	0.049	0.049	0.052	0.048	
		0.10	0.095	0.097	0.095	0.094	0.096	0.098	0.098	0.099	0.102	0.099	
	200	0.01	0.009	0.009	0.010	0.011	0.010	0.007	0.007	0.007	0.008	0.007	
		0.05	0.050	0.051	0.050	0.049	0.050	0.042	0.041	0.044	0.044	0.044	
		0.10	0.102	0.101	0.101	0.097	0.102	0.091	0.092	0.093	0.093	0.092	
	500	0.01	0.009	0.010	0.009	0.009	0.009	0.009	0.009	0.008	0.009	0.007	
		0.05	0.049	0.050	0.049	0.050	0.048	0.046	0.047	0.047	0.050	0.047	
		0.10	0.098	0.097	0.097	0.097	0.098	0.099	0.099	0.098	0.095	0.097	
Power													
AR(1): $\rho_1=0.25$	50	0.01	0.054	0.057	0.036	0.017	0.038	0.057	0.058	0.038	0.018	0.040	
		0.05	0.172	0.174	0.122	0.080	0.131	0.166	0.169	0.124	0.088	0.131	
		0.10	0.265	0.269	0.200	0.155	0.208	0.265	0.268	0.202	0.158	0.212	
	100	0.01	0.195	0.205	0.149	0.142	0.139	0.181	0.191	0.129	0.128	0.127	
		0.05	0.401	0.408	0.323	0.318	0.313	0.379	0.388	0.309	0.303	0.300	
		0.10	0.527	0.529	0.433	0.438	0.426	0.507	0.515	0.427	0.429	0.417	
	200	0.01	0.520	0.551	0.500	0.583	0.454	0.546	0.560	0.508	0.580	0.465	
		0.05	0.770	0.782	0.725	0.772	0.694	0.775	0.783	0.725	0.771	0.696	
		0.10	0.857	0.860	0.819	0.847	0.796	0.855	0.859	0.821	0.850	0.798	
	500	0.01	0.963	0.974	0.982	0.991	0.973	0.979	0.982	0.983	0.992	0.974	
		0.05	0.995	0.996	0.996	0.998	0.994	0.997	0.997	0.997	0.998	0.995	
		0.10	0.998	0.999	0.999	0.999	0.998	0.999	0.999	0.999	0.999	0.998	
AR(2): $\rho_1=0.05$ $\rho_2=0.10$	50	0.01	0.025	0.025	0.007	0.007	0.008	0.023	0.023	0.008	0.007	0.008	
		0.05	0.091	0.091	0.039	0.037	0.040	0.084	0.085	0.040	0.039	0.039	
		0.10	0.159	0.160	0.081	0.075	0.082	0.155	0.156	0.081	0.082	0.079	
	100	0.01	0.036	0.038	0.016	0.011	0.016	0.037	0.039	0.014	0.012	0.014	
		0.05	0.124	0.125	0.063	0.058	0.064	0.122	0.125	0.062	0.058	0.061	
		0.10	0.205	0.207	0.118	0.112	0.117	0.204	0.208	0.116	0.109	0.117	
	200	0.01	0.070	0.072	0.046	0.034	0.044	0.073	0.079	0.045	0.036	0.044	
		0.05	0.214	0.215	0.137	0.121	0.137	0.209	0.216	0.131	0.117	0.131	
		0.10	0.317	0.320	0.216	0.205	0.218	0.314	0.320	0.210	0.194	0.211	
	500	0.01	0.271	0.280	0.221	0.186	0.217	0.273	0.280	0.221	0.196	0.214	
		0.05	0.497	0.503	0.425	0.399	0.417	0.487	0.488	0.425	0.396	0.418	
		0.10	0.630	0.633	0.552	0.521	0.545	0.615	0.618	0.547	0.520	0.540	

Table S6: Sizes and powers of the new tests and three competing tests estimated via bootstrap sampling when $d = 4$, the error term is either normal or mixture normal, and the regressors are $X3$ (continued).

DGP	T	α	Normal errors					Mixture normal errors				
			NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p
AR(2): $\rho_1=0.05$ $\rho_2=0.20$	50	0.01	0.049	0.053	0.007	0.008	0.006	0.068	0.071	0.008	0.010	0.007
		0.05	0.153	0.154	0.039	0.039	0.039	0.191	0.191	0.042	0.042	0.041
		0.10	0.246	0.248	0.084	0.084	0.083	0.285	0.287	0.084	0.089	0.085
	100	0.01	0.145	0.145	0.052	0.047	0.051	0.138	0.140	0.045	0.041	0.043
		0.05	0.322	0.324	0.160	0.154	0.158	0.315	0.322	0.148	0.139	0.146
		0.10	0.445	0.449	0.253	0.240	0.253	0.445	0.446	0.234	0.224	0.234
	200	0.01	0.420	0.426	0.267	0.283	0.246	0.366	0.381	0.251	0.257	0.234
		0.05	0.625	0.631	0.472	0.488	0.451	0.617	0.626	0.468	0.470	0.450
		0.10	0.737	0.743	0.597	0.601	0.577	0.732	0.737	0.591	0.598	0.578
	500	0.01	0.902	0.906	0.877	0.898	0.846	0.889	0.903	0.866	0.891	0.842
		0.05	0.971	0.974	0.959	0.965	0.950	0.970	0.973	0.956	0.962	0.945
		0.10	0.987	0.988	0.979	0.982	0.975	0.986	0.986	0.977	0.982	0.971
AR(3): $\rho_1=0.10$ $\rho_2=0.10$ $\rho_3=0.10$	50	0.01	0.036	0.036	0.007	0.007	0.007	0.038	0.038	0.007	0.006	0.007
		0.05	0.118	0.117	0.039	0.036	0.039	0.115	0.114	0.035	0.035	0.035
		0.10	0.200	0.201	0.075	0.074	0.077	0.188	0.189	0.077	0.074	0.077
	100	0.01	0.089	0.095	0.039	0.025	0.039	0.076	0.080	0.032	0.018	0.034
		0.05	0.226	0.229	0.113	0.092	0.115	0.213	0.220	0.100	0.082	0.103
		0.10	0.329	0.336	0.186	0.164	0.189	0.321	0.323	0.168	0.149	0.171
	200	0.01	0.285	0.290	0.197	0.126	0.202	0.265	0.274	0.194	0.108	0.201
		0.05	0.490	0.492	0.379	0.292	0.388	0.484	0.487	0.364	0.282	0.370
		0.10	0.607	0.612	0.493	0.422	0.500	0.600	0.607	0.483	0.405	0.485
	500	0.01	0.783	0.802	0.783	0.587	0.785	0.781	0.808	0.778	0.597	0.790
		0.05	0.911	0.915	0.896	0.808	0.901	0.919	0.923	0.898	0.813	0.903
		0.10	0.949	0.951	0.934	0.887	0.937	0.954	0.955	0.936	0.891	0.937
AR(3): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.20$	50	0.01	0.036	0.037	0.024	0.012	0.026	0.032	0.032	0.025	0.012	0.026
		0.05	0.123	0.123	0.086	0.057	0.086	0.127	0.128	0.091	0.061	0.094
		0.10	0.204	0.204	0.141	0.111	0.145	0.214	0.215	0.151	0.116	0.155
	100	0.01	0.114	0.116	0.057	0.045	0.057	0.091	0.095	0.048	0.045	0.048
		0.05	0.274	0.277	0.168	0.161	0.168	0.258	0.264	0.161	0.149	0.156
		0.10	0.389	0.393	0.270	0.253	0.269	0.385	0.387	0.262	0.241	0.255
	200	0.01	0.373	0.376	0.274	0.284	0.252	0.340	0.356	0.285	0.294	0.264
		0.05	0.601	0.609	0.501	0.525	0.478	0.605	0.612	0.504	0.512	0.479
		0.10	0.725	0.727	0.632	0.640	0.608	0.717	0.718	0.626	0.632	0.604
	500	0.01	0.907	0.918	0.895	0.908	0.872	0.903	0.914	0.900	0.912	0.881
		0.05	0.975	0.977	0.968	0.972	0.959	0.974	0.976	0.969	0.973	0.963
		0.10	0.988	0.989	0.984	0.987	0.980	0.990	0.990	0.986	0.986	0.980

Table S6: Sizes and powers of the new tests and three competing tests estimated via bootstrap sampling when $d = 4$, the error term is either normal or mixture normal, and the regressors are $X3$ (continued).

DGP	T	α	Normal errors					Mixture normal errors					
			NRR	MCMC	Q	Min p	Prod p	NRR	MCMC	Q	Min p	Prod p	
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.15$ $\rho_4=0.15$	50	0.01	0.063	0.063	0.009	0.009	0.009	0.058	0.058	0.009	0.009	0.010	
		0.05	0.174	0.174	0.044	0.043	0.043	0.175	0.175	0.045	0.039	0.045	
		0.10	0.273	0.270	0.087	0.084	0.088	0.270	0.269	0.087	0.081	0.087	
	100	0.01	0.01	0.214	0.218	0.070	0.039	0.070	0.215	0.217	0.064	0.035	0.067
			0.05	0.411	0.416	0.177	0.136	0.178	0.399	0.400	0.169	0.128	0.173
		0.10	0.01	0.530	0.529	0.269	0.225	0.270	0.511	0.512	0.261	0.215	0.261
			0.05	0.575	0.590	0.449	0.295	0.458	0.602	0.614	0.466	0.303	0.473
		200	0.01	0.788	0.794	0.650	0.540	0.654	0.780	0.785	0.652	0.540	0.656
			0.05	0.865	0.867	0.744	0.669	0.747	0.853	0.854	0.744	0.668	0.744
	500	0.01	0.983	0.987	0.981	0.932	0.982	0.986	0.989	0.983	0.936	0.983	
		0.05	0.997	0.997	0.995	0.985	0.995	0.998	0.998	0.995	0.984	0.995	
		0.10	0.999	0.999	0.998	0.993	0.998	0.999	0.999	0.998	0.994	0.997	
AR(4): $\rho_1=0.05$ $\rho_2=0.10$ $\rho_3=0.05$ $\rho_4=0.10$	50	0.01	0.036	0.036	0.005	0.006	0.005	0.034	0.034	0.005	0.006	0.005	
		0.05	0.115	0.114	0.028	0.030	0.028	0.111	0.112	0.028	0.028	0.028	
		0.10	0.194	0.193	0.061	0.063	0.062	0.181	0.181	0.058	0.060	0.059	
	100	0.01	0.01	0.077	0.080	0.020	0.015	0.021	0.077	0.079	0.020	0.013	0.020
			0.05	0.212	0.215	0.072	0.062	0.073	0.201	0.206	0.072	0.058	0.073
		0.10	0.01	0.318	0.319	0.127	0.116	0.129	0.303	0.306	0.125	0.111	0.128
			0.05	0.223	0.233	0.143	0.081	0.153	0.207	0.220	0.132	0.077	0.132
		200	0.01	0.442	0.447	0.288	0.221	0.293	0.419	0.424	0.269	0.207	0.272
			0.05	0.557	0.562	0.386	0.325	0.389	0.534	0.538	0.373	0.315	0.374
	500	0.01	0.681	0.711	0.658	0.462	0.670	0.669	0.684	0.637	0.445	0.638	
		0.05	0.861	0.865	0.811	0.700	0.813	0.849	0.857	0.803	0.698	0.806	
		0.10	0.914	0.915	0.868	0.802	0.871	0.912	0.916	0.869	0.800	0.873	
AR(4): $\rho_1=0.05$ $\rho_2=0.05$ $\rho_3=0.05$ $\rho_3=0.05$	50	0.01	0.021	0.021	0.008	0.007	0.007	0.021	0.021	0.008	0.007	0.007	
		0.05	0.083	0.082	0.036	0.035	0.036	0.083	0.082	0.036	0.035	0.036	
		0.10	0.144	0.143	0.071	0.069	0.073	0.144	0.143	0.071	0.069	0.073	
	100	0.01	0.01	0.032	0.033	0.010	0.008	0.010	0.032	0.033	0.010	0.008	0.010
			0.05	0.108	0.112	0.043	0.043	0.044	0.108	0.112	0.043	0.043	0.044
		0.10	0.01	0.184	0.185	0.088	0.087	0.089	0.184	0.185	0.088	0.087	0.089
			0.05	0.062	0.068	0.038	0.023	0.039	0.062	0.068	0.038	0.023	0.039
		200	0.01	0.180	0.183	0.105	0.089	0.107	0.180	0.183	0.105	0.089	0.107
			0.05	0.272	0.274	0.173	0.154	0.176	0.272	0.274	0.173	0.154	0.176
	500	0.01	0.227	0.231	0.190	0.106	0.199	0.227	0.231	0.190	0.106	0.199	
		0.05	0.428	0.441	0.364	0.276	0.373	0.428	0.441	0.364	0.276	0.373	
		0.10	0.561	0.569	0.479	0.402	0.485	0.561	0.569	0.479	0.402	0.485	

Table S7: Sizes of the new tests estimated through parametric sampling when the regressor is X_1 and $d = 20$ and 50 .

size	m	$d = 20$			$d = 50$		
		0.01	0.05	0.10	0.01	0.05	0.10
100	10,000	0.010	0.053	0.104	0.010	0.055	0.104
	20,000	0.011	0.051	0.105	0.011	0.051	0.101
	30,000	0.010	0.050	0.103	0.009	0.050	0.100
	50,000	0.009	0.051	0.104	0.009	0.049	0.099
	75,000	0.010	0.052	0.103	0.009	0.048	0.098
	100,000	0.010	0.052	0.103	0.009	0.049	0.097
200	10,000	0.012	0.056	0.102	0.010	0.054	0.101
	20,000	0.011	0.052	0.099	0.011	0.051	0.100
	30,000	0.011	0.051	0.099	0.010	0.050	0.098
	50,000	0.011	0.048	0.100	0.010	0.050	0.099
	75,000	0.011	0.050	0.100	0.011	0.051	0.101
	100,000	0.011	0.050	0.101	0.011	0.050	0.100
300	10,000	0.012	0.053	0.104	0.011	0.052	0.100
	20,000	0.011	0.052	0.100	0.011	0.051	0.101
	30,000	0.010	0.054	0.104	0.010	0.049	0.101
	50,000	0.011	0.052	0.102	0.010	0.050	0.102
	75,000	0.011	0.051	0.102	0.010	0.050	0.102
	100,000	0.011	0.052	0.102	0.010	0.050	0.101
500	10,000	0.010	0.046	0.090	0.011	0.049	0.102
	20,000	0.009	0.045	0.095	0.011	0.051	0.100
	30,000	0.009	0.045	0.097	0.010	0.051	0.103
	50,000	0.009	0.045	0.095	0.009	0.050	0.102
	75,000	0.009	0.046	0.095	0.011	0.050	0.100
	100,000	0.009	0.046	0.093	0.010	0.050	0.100
1000	10,000	0.010	0.049	0.099	0.009	0.045	0.092
	20,000	0.011	0.048	0.095	0.009	0.045	0.096
	30,000	0.010	0.048	0.094	0.009	0.045	0.094
	50,000	0.010	0.048	0.096	0.010	0.047	0.094
	75,000	0.010	0.047	0.094	0.010	0.046	0.095
	100,000	0.010	0.047	0.097	0.009	0.045	0.097