Exact point-optimal sign-based tests for predictive linear and nonlinear regressions

Jean-Marie Dufour*	Kaveh Salehzadeh Nobari†	Abderrahim Taamouti‡

McGill University	Durham University	Durham University

November 17, 2020

*William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ).Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

†Department of Economics and Finance, Durham University. Address and email: Department of Economics and Finance, Durham University Business School, Mill Hill Lane, Durham, DH1 3LB.

‡Department of Economics and Finance, Durham University. Address and email: Department of Economics and Finance, Durham University Business School, Mill Hill Lane, Durham, DH1 3LB.

Exact point-optimal sign-based tests for predictive linear and nonlinear regressions

ABSTRACT

Predictors of stock returns (e.g. dividend-price ratio, earnings-price ratio, etc.) are often highly persistent with innovations that are correlated with the disturbances in the predictive regression of returns, which as known leads to invalid inference using conventional tests. We propose point-optimal sign-based tests in the context of linear and nonlinear models that are valid in the presence of stochastic regressors. The proposed tests are exact, distribution-free, and robust against heteroskedasticity of unknown form. Further, they may be inverted to build confidence regions for the parameters of the regression function. Point-optimal tests maximize the power at a predetermined point in the alternative hypothesis parameter space, which in our case is unknown. Therefore, we propose an adaptive approach based on the split-sample technique to shift the power function close to that of the power envelope (i.e. maximum obtainable power for a given testing problem). We then present a Monte Carlo study to assess the performance of the proposed "quasi"-point-optimal sign test by comparing its size and power to those of certain existing tests which are supposed to be robust against heteroskedasticity. The results show that our procedures are superior. Finally, an empirical application is considered to illustrate the usefulness of our proposed tests for testing the predictability of stock returns.

Journal of Economic Classification: C01, C13, C15, C22, C52

Keywords: Stochastic regressors; stock return predictability; valuation ratios; persistency; sign test; point-optimal test; nonlinear model; heteroskedasticity; exact inference; distribution-free; split-sample; adaptive method; projection technique; numerical optimization.

Contents

1 Introduction

Numerous studies investigate the predictability of financial and economic variables using the past values of one or more predictors. The most commonly encountered examples of such studies concern the predictability of stock returns using the lag of certain fundamental variables, such as the dividend-yield, earnings-price ratio or interest rates [see Campbell and Shiller (1988), Fama and French (1988), Campbell and Yogo (2006), Campbell and Thompson (2008), and Golez and Koudijs (2018), among others]. Predictability in this context is generally assessed using the OLS regression of the returns against the said predictors and tested with conventional T-type tests. However, the predictors that are often considered in these studies are known to be highly persistent with innovations that are correlated with the disturbances in the predictive regression of the returns. In such situations, we know that the OLS estimator of the coefficients, although consistent, will be biased. As a result of this bias, in finite samples the T-statistic will have a nonstandard distribution which leads to invalid inference [see Mankiw and Shapiro (1986), Banerjee et al. (1993) and Stambaugh (1999) among others]. In this paper, we address this issue by deriving point-optimal sign-based tests (hereafter POS-based tests) in the context of linear and non-linear predictive regressions that are distribution-free, robust against heteroskedasticity of unknown form and which allow for serial (non-linear) dependence provided that the residual process has zero median conditional on the explanatory variables and its own past. This assumption allows the signs to be i.i.d under the null hypothesis of orthogonality according to a known distribution, despite the fact that the variables to which the indicator functions are applied are dependent [see Coudin and Dufour (2009)].

Nelson and Kim (1993) reduce the small-sample bias using bootstrap simulations and Stambaugh (1999) shows that in the case of stationary regressors the said bias can be corrected. However, in later studies Phillips and Lee (2013) and Phillips (2014) show this to be infeasible in the presence of predictors that exhibit local-to-unity, unit-root or explosive persistency. Therefore, many inference procedures in this context address the issue of size distortions by considering local-to-unity asymptotics, where the key predictor variable is assumed to contain a unit root [Lewellen (2004)], or can be modeled as having a local-to-unit root [Elliott and Stock (1994), Torous et al. (2004), and Campbell and Yogo (2006), among others]. Notable studies under the local-to-unity dynamics employ an array of procedures, such as Bonferroni corrections [e.g. Cavanagh et al. (1995) and Campbell and Yogo (2006)], a conditional likelihood based approach [e.g. Jansson and Moreira (2006)], as well as the nearly optimal tests proposed by Elliott et al. (2015). In more recent work, Kostakis et al. (2015) and Phillips and Lee (2016) expand on the predictability literature by utilizing an extension of the instrumental variable procedure suggested by Phillips et al. (2009) to generalize inference to multivariate regressors with stationary, local-to-unity and explosive persistency. The contribution of the POS-based tests proposed in our study is twofold: firstly, as our tests are distribution-free, they are valid in the presence of regressors with general persistency in finite samples and do not suffer from the discontinuity that is commonly observed in the limiting distribution of conventional test statistics between stationary, local-to-unity and explosive autoregressions. Secondly, our tests possess the greatest power among certain parametric and non-parametric tests that are often encountered in practice and can easily be extended to multivariate testing problems.

In a recent study, Dufour and Taamouti (2010) propose simple point-optimal sign-based tests in the context of linear and non-linear regression models, which are valid under non-normality and heteroskedasticity of unknown form, provided the errors have zero median conditional on the explanatory variables. These tests are exact, distribution-free, and robust against heteroskedasticity of unknown form, and may be inverted to build confidence regions for the vector of unknown parameters. This work, however, is developed under the assumption that the predictors are fixed; thus, these tests are not applicable in the presence of stochastic regressors. We extend the above tests to the case where the predictors can be fixed or stochastic. The main motivation is to build point-optimal sign-based tests for linear and non-linear predictability of stock returns that retain the advantages of the POS-based tests proposed by Dufour and Taamouti (2010).

To extend the previous work of Dufour and Taamouti (2010), we recognize that under the alternative hypothesis the signs are no longer necessarily independent and the test-statistic now depends on the joint distribution of the signs, which is computationally infeasible. Therefore, an additional assumption on the dependence structure of the process of signs is needed to obtain *feasible* test statistics. In particular, we assume that this process is a Markov process of finite-order. By construction, our POS-based tests control the size for any given sample. Under the null hypothesis of unpredictability, the tests are valid even in the presence of the bias problem pointed out by Mankiw and Shapiro (1986) and Stambaugh (1985, 1999), which affects the classical testing procedure for stock returns predictability. In addition, our tests are model-free and robust against heteroskedasticity of unknown form. The tests are point-optimal tests, which are useful in a number of ways and are particularly attractive when testing one financial theory against another. An important feature of these tests stems from the fact that they trace out the power envelope, i.e. the maximum achievable power for a given testing problem, which may be used as a benchmark against which other testing procedures can be evaluated. Finally, our tests may be inverted to build confidence regions for the parameters of the regression function.

As point-optimal tests maximize the power at a nominated point in the alternative hypothesis parameter space, a practical problem concerns finding an alternative at which the power curve of the POS-based test is close to that of the power envelope. Following Dufour and Torrès (1998), Dufour and Jasiak (2001) and Dufour and Taamouti (2010), we propose an adaptive approach based on the split-sample technique to choose the alternative hypothesis. This procedure consists of splitting the sample into two independent sub-samples, where the first part is used to estimate the alternative hypothesis and the second part to compute the POS-based test statistic [see Dufour and Iglesias (2008)]. In a simulations exercise, Dufour and Taamouti (2010) find that using the first 10% of the sample to estimate the alternative and the rest to compute the test statistic, achieves a power that traces out the power envelope. We present a Monte Carlo study to assess the performance of the proposed "quasi"-POS-based tests by comparing its size and power to certain existing tests that are intended to be robust against heteroskedasticity. We show the superiority of our procedures in the presence of nearly integrated regressors and under different distributional assumptions and forms of heteroskedasticity..

The plan of the paper is as follows: Section 2 provides exact POS-based tests in the context of linear and non-linear predictive regressions. Section 3 discusses the selection approach for the alternative hypothesis to compute the POS-based test statistic. Section 4, discusses the construction of POS-based confidence regions using the projection techniques. Section 5, presents a Monte Carlo study to assess the performance of the POS-based tests by comparing their size and power to those of certain popular tests. Section 6 is devoted to an empirical application, and finally, the paper is concluded in Section 7. Proofs are presented in Appendix 8.

2 POS tests in linear and non-linear regression models

In this Section, we follow the structure of Dufour and Taamouti (2010) to derive POS-based tests in the context of linear and non-linear regression models. However, in our study stochastic regressors may as well be considered. First, we divert our attention to the problem of testing whether the conditional median of a vector of observations is zero against a linear regression alternative. This is is later generalized to test whether the coefficients of a possibly non-linear median regression function have a given value against an alternative non-linear median regression. Although the former problem is a special case of the latter, for the simplicity of exposition the linear regression model is considered first.

2.1 Testing independence (zero coefficients) hypothesis in linear regressions

Consider a stochastic process $Z = \{Z_t = (y_t, x_t') : \Omega \to \mathbb{R}^{(k+1)} : t = 0, 1, \cdots\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. Let $\{Z_t, \mathcal{F}_t\}_{t=0,1,\ldots}$ be an adapted stochastic sequence, such that $\mathcal{F}_t$ is a σ-field in Ω, $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s < t$, $\sigma(Z_0, \cdots, Z_t) \subset \mathcal{F}_t$, where $\sigma(Z_0, \cdots, Z_t)$ is the σ-field generated by $Z_0, \cdots, Z_t$. Suppose that y_t can linearly be explained by a vector variable x_t

$$y_t = \beta' x_{t-1} + \varepsilon_t, \quad t = 1, ..., n, \tag{1}$$

where x_{t-1} is an $(k+1) \times 1$ vector of stochastic explanatory variables, say $x_{t-1} = [1, x_{1,t-1}, ..., x_{k,t-1}]'$, $\beta \in \mathbb{R}^{(k+1)}$ is an unknown vector of parameters with $\beta = [\beta_0, \beta_1, ..., \beta_k]'$ and

$$\varepsilon_t \mid X \sim F_t(. \mid X)$$

where $F_t(. \mid X)$ is an unknown conditional distribution function and X=$[x_0, \cdots, x_{n-1}]'$ is an $n \times (k+1)$ matrix.

In the context of general forms of (non-linear) dependence, an assumption that is commonly made on the error terms $\{\varepsilon_t, t = 1 \cdots, n\}$ is that the error process is a martingale difference sequence (MDS hereafter) with respect to $\mathcal{F}_t = \sigma(Z_0, \cdots, Z_t)$ for $t = 0, 1, \cdots,$ - i.e. $\mathbb{E}\{\varepsilon_t \mid \mathcal{F}_{t-1}\} = 0, \quad \forall t \geq 1$. As the latter assumption relies on the first moment of the residuals, we follow Coudin and Dufour (2009) by departing from this assumption and considering the median as an alternative measure of central tendency.

This implies imposing a median-based analogue of the MDS on the error process - namely we suppose that ε_t is a strict conditional mediangale

$$P[\varepsilon_t > 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = P[\varepsilon_t < 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = \frac{1}{2}, \tag{2}$$

with

$$\boldsymbol{\varepsilon}_0 = \{\emptyset\}, \quad \boldsymbol{\varepsilon}_{t-1} = \{\varepsilon_1, \cdots, \varepsilon_{t-1}\}, \quad \text{for} \quad t \geq 2$$

Note (2) entails that $\varepsilon_t \mid X$ has no mass at zero for all t, which is only true if $\varepsilon_t \mid X$ is a continuous variable. Model (1) in conjunction with assumption (2) allows the error terms to possess asymmetric, heteroskedastic and serially (non-linear) dependent distributions, so long as the conditional medians are zero. Assumption 2 allows for many dependent schemes, such as those of the form $\varepsilon_1 = \sigma_1(x_0, \cdots, x_{t-2})\epsilon_1$, $\varepsilon_t = \sigma_t(x_0, \cdots, x_{t-2}, \varepsilon_1, \cdots, \varepsilon_{t-1})\epsilon_t$, $t = 2, \cdots, n$, where $\epsilon_1, \cdots, \epsilon_n$ are independent with a zero median. In time-series context this includes models such as ARCH, GARCH or stochastic volatility with non-Gaussian errors. Furthermore, in the mediangale framework the disturbances need not be second order stationary.

We wish to test the null hypothesis

$$H_0 : \beta = 0 \tag{3}$$

against the alternative H_1

$$H_1 : \beta = \beta_1. \tag{4}$$

We define the following vector of signs

$$U(n) = (s(y_1), ..., s(y_n))',$$

where, for $1 \leq t \leq n$,

$$s(y_t) = \begin{cases} 1, \; if \;\; y_t \geq 0 \\ 0, \; if \;\; y_t < 0 \end{cases}.$$

The test is Neyman-Pearson type test based on signs [see Lehmann and Romano (2006)] which maximize the power function under the constraint $P\,[\text{reject } H_0 \mid H_0] \leq \alpha$. The idea is to build point-optimal sign-based tests to test the null hypothesis (3) against the alternative hypothesis (4). To do so, we first define the likelihood function of sample in terms of signs $s(y_1), ..., s(y_n)$ conditional on X

$$L\left(U(n), \beta\right) = P\left[s(y_1) = s_1, ..., s(y_n) = s_n \mid X\right] = \prod_{t=1}^{n} P\left[s(y_t) = s_t \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right],$$

with

$$\underset{\sim}{\mathrm{S}}_0 = \{\emptyset\}\,, \;\; \underset{\sim}{\mathrm{S}}_{t-1} = \{s(y_1) = s_1, ..., s(y_{t-1}) = s_{t-1}\}\,, \text{ for } t \geq 2,$$

and

$$P\left[s(y_1) = s_1 \mid \underset{\sim}{S}_0, X\right] = P\left[s(y_1) = s_1 \mid X\right],$$

where each s_t, for $1 \leq t \leq n$, takes two possible values 0 and 1.

As the error terms satisfy the strict conditional mediangale assumption (2), the distribution of the signs $s(\varepsilon_1), \cdots, s(\varepsilon_n)$, and in turn $s(y_1), \cdots, s(y_n)$ under the null hypothesis of orthogonality is specified and are mutually independent [see Coudin and Dufour (2009)].

Theorem 1 *Under model (1) and assumption (2), the variables $s(\varepsilon_1), \cdots, s(\varepsilon_n)$ are i.i.d conditional on X, according to the distribution*

$$P[s(\varepsilon_1) = 1 \mid X] = P[s(\varepsilon_1) = 0 \mid X] = \frac{1}{2}, \quad t = 1, \cdots, n$$

This result holds true iff for any combination of $t = 1, \cdots, n$ there is a permutation $\pi : i \to j$ such that the mediangale assumption holds for j. Then the signs $s(\varepsilon_1), \cdots, s(\varepsilon_n)$ are i.i.d.
Proof: *See Appendix.*

A sign-based test for testing the null hypothesis (3) against the alternative hypothesis (4) is given by the following proposition.

Proposition 1 *Under assumptions (2) and (1), let H_0 and H_1 be defined by (3) - (4),*

$$SL_n(\beta_1) = \sum_{t=1}^{n} a_t(\beta_1) s(y_t),$$

where, for $t = 1, ..., n$,

$$a_t(\beta_1) = \ln \left\{ \frac{P\left[y_t \geq 0 \mid \underset{\sim}{S}_{t-1}, X\right]}{P\left[y_t < 0 \mid \underset{\sim}{S}_{t-1}, X\right]} \right\}, \tag{5}$$

and suppose the constant $c_1(\beta_1)$ satisfies $P\left[\sum_{t=1}^{n} a_t(\beta_1) s(y_t) > c_1(\beta_1)\right] = \alpha$ under H_0, with $0 < \alpha < 1$. Then the test that rejects H_0 when

$$SL_n(\beta_1) > c_1(\beta_1) \tag{6}$$

is most powerful for testing H_0 against H_1 among level-α tests based on the signs $\left(s(y_1), ..., s(y_n)\right)'$.
Proof: *See Appendix.*

Notice that the calculation of the test statistic $SL_n(\beta_1)$ depends on the weights $a_t(\beta_1)$, which in turn depends on the calculation of the conditional probabilities $P\left[y_t \geq 0 \mid \underset{\sim}{S}_{t-1}, X\right]$ and $P\left[y_t < 0 \mid \underset{\sim}{S}_{t-1}, X\right]$. The latter terms are not easy to compute and involves the distribution of the joint process of signs $\left(s(y_1), ..., s(y_n)\right)'$ conditional on X, which is unknown. An alternative way to compute the terms $P\left[y_t \geq 0 \mid \underset{\sim}{S}_{t-1}, X\right]$ and $P\left[y_t < 0 \mid \underset{\sim}{S}_{t-1}, X\right]$ is to use simulations, however again this will be time consuming as it requires to simulate the joint distribution of the process of signs $\left(s(y_1), ..., s(y_n)\right)'$, which depends on the sample size n. For all these reasons and to make the test statistic $SL_n(\beta_1)$ feasible, we make the following assumption.

Assumption A1: *Let $\{y_t, t=0,1,\cdots\}$ follow a Markov process of order one. Then under the alternative hypothesis, the sign process $\{s(y_t)\}_{t=0}^{\infty}$ follows a Markov process of the same order.*

Proof: *See Appendix.*

As the mediangale assumption allows for non-linear serial dependence, testing assumption **A1** by considering linear correlation is inappropriate. One approach involves fitting copula models, which provides the means of separating the marginal distributions of the process from their respective dependence structure. The latter stems from Sklar (1959), which decomposes the joint conditional distribution function of $\boldsymbol{Y} = [y_1, \cdots, y_n]'$ conditional on X as

$$\boldsymbol{Y} \mid X \sim H(. \mid X) = C\left(F_1(. \mid X), \cdots, F_n(. \mid X)\right),$$

where $F_t(. \mid X)$ for $t = 1, \cdots, n$ are uniformly distributed marginals - i.e. $F_t(. \mid X) := u_t \sim U[0,1]$. Note that the elements of $\boldsymbol{Y}$ are uncorrelated, yet exhibit serial dependence which is captured by the copula $C(.)$. The issue with specifying a copula for $\boldsymbol{Y}$ is that the no serial correlation assumption implies an identity correlation matrix. As a result, in the literature, the means of allowing for non-linear serial dependence for processes which are linearly unrelated is often accompanied by assuming that $\boldsymbol{Y}$ conditional on X is distributed as a multivariate Student's t distribution - i.e. $\boldsymbol{Y} \mid X \sim t_\nu(0, I)$, where I is an identity matrix. When I is imposed on the multivariate Student's t distribution, the conditional joint distribution of $\boldsymbol{Y}$ does not factorize into the product of its marginals. Alternatively, we may consider the "jointly symmetric"copulae proposed by Oh and Patton (2016), where the latter can be constructed with any given (possibly asymmetric) copula family. In addition, when they are combined with symmetric marginals, they ensure an identity correlation matrix. A "jointly symmetric"copula is defined as follows

Definition 1 *The n dimensional copula C^{JS}, is jointly symmetric:*

$$C^{JS}\left(u_1, \cdots, u_n\right) = \frac{1}{2^n} \sum_{k_1=0}^{2} \cdots \sum_{k_n=0}^{2} (-1)^R C(\tilde{u}_1, \cdots, \tilde{u}_i, \cdots, \tilde{u}_n)$$

$$\textit{where} \quad R = \sum_{i=1}^{n} \mathbf{1}\{k_i = 2\}, \quad \textit{and} \quad \tilde{u}_i = \begin{cases} 1, & k_i = 0 \\ u_i, & k_i = 1 \\ 1 - u_i, & k_i = 2 \end{cases}$$

The general idea is that the average of mirror image rotations of a possibly asymmetric copula along each axis generates a jointly symmetric copula [see Oh and Patton (2016). For instance, the marginals can be assumed to possess standard normal distributions, while the nonlinear dependency is modeled using the jointly symmetric copulae. The Markovian assumption can then be tested by considering in turn, the independent, bivariate, trivariate and higher order multivariate copulae (or multivariate Student's t distributions of varying orders), where the model with the lowest Akaike Information Criterion (AIC hereafter) is then chosen. However, in an extensive simulation analysis, we observe that the order of the Markovianity does not have a significant impact on the power of the test, and as such the assumption that under the alternative hypothesis $\{y_t : t = 1, \cdots, n\}$ and in turn $\{s(y_t) : t = 1, \cdots, n\}$ follow a Markov process of order one is sufficient for testing the null hypothesis of orthogonality.

Now, under assumption **A1**, the probability terms $P\left[y_t \geq 0 \mid \underset{\sim}{S}_{t-1}, X\right]$ and $P\left[y_t < 0 \mid \underset{\sim}{S}_{t-1}, X\right]$ in the weight function $a_t(\beta_1)$ can be written as follows:

$$\begin{cases} P\left[y_t \geq 0 \mid \underset{\sim}{S}_{t-1}, X\right] = P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]^{s(y_{t-1})} P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]^{1-s(y_{t-1})}, \\ \\ P\left[y_t < 0 \mid \underset{\sim}{S}_{t-1}, X\right] = P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right]^{s(y_{t-1})} P\left[y_t < 0 \mid y_{t-1} < 0, X\right]^{1-s(y_{t-1})}. \end{cases}$$

The use of the above expressions of $P\left[y_t \geq 0 \mid \underset{\sim}{S}_{t-1}, X\right]$ and $P\left[y_t < 0 \mid \underset{\sim}{S}_{t-1}, X\right]$ simplifies a lot the calculation of the test statistic $SL_n(\beta_1)$. We have the following result.

Corollary 1 *Under assumptions (2) and (1), let H_0 and H_1 be defined by (3) - (4),*

$$\tilde{S}L_n(\beta_1) = \sum_{t=1}^{n} \tilde{a}_t(\beta_1) s(y_t) + \sum_{t=1}^{n} \tilde{b}_t(\beta_1) s(y_t) s(y_{t-1}),$$

where

$$\tilde{a}_1(\beta_1) = \ln\left\{\frac{1 - P\left[\varepsilon_1 < -\beta_1' x_0 \mid X\right]}{P\left[\varepsilon_1 < -\beta_1' x_0 \mid X\right]}\right\}, \quad \tilde{b}_1(\beta_1) = 0$$

and for $t = 2, ..., n$,

$$\tilde{a}_t(\beta_1) = \ln\left\{\frac{1 - \frac{P\left[\varepsilon_t < -\beta_1' x_{t-1},\ \varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}{P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}}{\frac{P\left[\varepsilon_t < -\beta_1' x_{t-1},\ \varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}{P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}}\right\}$$

$$\tilde{b}_t(\beta_1) = \ln\left\{\frac{1 - \left(\frac{P\left[\varepsilon_t < -\beta_1' x_{t-1} \mid X\right]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]} - \frac{P\left[\ \varepsilon_{t-1} < -\beta_1' x_{t-2},\ \varepsilon_t < -\beta_1' x_{t-1} \mid X\right]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}\right)}{\frac{P\left[\varepsilon_t < -\beta_1' x_{t-1} \mid X\right]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]} - \frac{P\left[\ \varepsilon_{t-1} < -\beta_1' x_{t-2},\ \varepsilon_t < -\beta_1' x_{t-1} \mid X\right]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}}\right\} - \ln\left\{\frac{1 - \frac{P\left[\varepsilon_t < -\beta_1' x_{t-1},\ \varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}{P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}}{\frac{P\left[\varepsilon_t < -\beta_1' x_{t-1},\ \varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}{P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}}\right\}$$

and suppose the constant $\tilde{c}_1(\beta_1)$ satisfies $P\left[\sum_{t=1}^{n} \tilde{a}_t(\beta_1)\ s(y_t) + \sum_{t=1}^{n} \tilde{b}_t(\beta_1)\ s(y_t) s(y_{t-1}) > \tilde{c}_1(\beta_1)\right] = \alpha$ under H_0, with $0 < \alpha < 1$. Then the test that rejects H_0 when

$$\tilde{S}L_n(\beta_1) > \tilde{c}_1(\beta_1) \tag{7}$$

is most powerful for testing H_0 against H_1 among level-α tests based on the signs $\left(s(y_1), ..., s(y_n)\right)'$.

Proof: *See Appendix.*

Now, the calculation of the test statistic $\tilde{S}L_n(\beta_1)$ depends on the univariate and bivariate distributions $P\left[\varepsilon_t < \cdot \mid X\right]$ and $P\left[\ \varepsilon_{t-1} < \cdot,\ \varepsilon_t < \cdot \mid X\right]$,

Observe that under the null hypothesis, the signs $s(y_1), ..., s(y_n)$ are i.i.d. according to a Bernoulli $Bi(1, 0.5)$. Thus, the distribution of the test statistic $\tilde{S}L_n(\beta_1)$ only depends on the known weights $\tilde{a}_t(\beta_1)$ and $\tilde{b}_t(\beta_1)$ and does not involve any nuisance parameter under the null hypothesis. Nonparametric assumption (2) implies that tests based on $\tilde{S}L_n(\beta_1)$, such as the test given by (7), are distribution-free and robust against heteroskedasticity of unknown form, and thus, a nonparametric *pivotal function*. Under the alternative hypothesis, however, the power function of the test depends on the form of the distributions $P\left[\varepsilon_t < . \mid X\right]$ and $P\left[\ \varepsilon_{t-1} < \cdot,\ \varepsilon_t < \cdot \mid X\right]$.

A special case is where $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{n-1}, \varepsilon_n$ are distributed according to $N(0, 1)$. As suggested before, since the form of the serial dependence of the errors is non-linear, we may calculate the bivariate probabilities using "jointly-symmetric"copulae by considering the Archimedean Frank, Clayton or Gumbel as the copula family [see Joe (2014)]. Alternatively, we may evaluate the bivariate probabilities $P\left[\ \varepsilon_{t-1} < \cdot,\ \varepsilon_t < \cdot \mid X\right]$ using a multivariate Student's t distribution by imposing the identity matrix I. Then the optimal test statistic takes the form

$$\tilde{S}L_n(\beta_1) = \sum_{t=1}^{n} \tilde{a}_t(\beta_1)s(y_t) + \sum_{t=1}^{n} \tilde{b}_t(\beta_1)s(y_t)s(y_{t-1}),$$

where

$$\tilde{a}_1(\beta_1) = \ln\left\{\frac{\Phi(\beta_1' x_0)}{1-\Phi(\beta_1' x_0)}\right\},\ \tilde{b}_1(\beta_1) = 0$$

and for $t = 2, ..., n$,

$$\tilde{a}_t(\beta_1) = \ln\left\{\frac{1-\frac{C^{JS}(\Phi(-\beta_1' x_{t-1}),\Phi(-\beta_1' x_{t-2}))}{1-\Phi\left(\beta_1' x_{t-2}\right)}}{\frac{C^{JS}(\Phi(-\beta_1' x_{t-1}),\Phi(-\beta_1' x_{t-2}))}{1-\Phi\left(\beta_1' x_{t-2}\right)}}\right\}$$

$$\tilde{b}_t(\beta_1) = \ln\left\{\frac{1-\left(\frac{1-\Phi(\beta_1' x_{t-1})}{\Phi\left(\beta_1' x_{t-2}\right)}-\frac{C^{JS}(\Phi(-\beta_1' x_{t-1}),\Phi(-\beta_1' x_{t-2}))}{\Phi\left(\beta_1' x_{t-2}\right)}\right)}{\frac{1-\Phi\left(\beta_1' x_{t-1}\right)}{\Phi\left(\beta_1' x_{t-2}\right)}-\frac{C^{JS}(\Phi(-\beta_1' x_{t-1}),\Phi(-\beta_1' x_{t-2}))}{\Phi\left(\beta_1' x_{t-2}\right)}}\right\} - \ln\left\{\frac{1-\frac{C^{JS}(\Phi(-\beta_1' x_{t-1}),\Phi(-\beta_1' x_{t-2}))}{1-\Phi\left(\beta_1' x_{t-2}\right)}}{\frac{C^{JS}(\Phi(-\beta_1' x_{t-1}),\Phi(-\beta_1' x_{t-2}))}{1-\Phi\left(\beta_1' x_{t-2}\right)}}\right\}$$

where $\Phi(.)$ is the standard normal distribution function and $C^{JS}(u_1, u_2)$ is the bivariate "jointly symmetric"copula with uniformly distributed margins.

The distribution of $\tilde{S}L_n(\beta_1)$, can be simulated under the null hypothesis and the relevant critical values can be evaluated to any degree of precision with a sufficient number of replications. Since the test statistic $\tilde{S}L_n(\beta_1)$ is a continuous random variable, its quantiles are easy to compute. To simulate the distribution of $\tilde{S}L_n(\beta_1)$, the following algorithm is implemented:

1. Compute the test statistic $\tilde{S}L_n(\beta_1)$ based on the observed data, say $\tilde{S}L_n^0(\beta_1)$;
2. Generate a sample $\{y_t\}_{t=1}^n$ of length n under the null H_0 and compute $\tilde{S}L_n^j(\beta_1)$ using that generated sample;
3. Choose B such that $\alpha(B+1)$ is an integer and repeat steps 1-2 B times;
4. Computer the $(1-\alpha)\%$ quantile, say $\tilde{c}_1(\beta_1)$, of the sequence $\{\tilde{S}L_n^j(\beta_1)\}_{j=1}^B$;
5. Reject the null hypothesis at level α if $\tilde{S}L_n^0(\beta_1) \geq c(\beta_1)$.

2.2 Testing general full coefficient hypotheses in non-linear regressions

We now consider a non-linear regression model

$$y_t = f(x_{t-1}, \beta) + \varepsilon_t,\ t = 1, ..., n, \tag{8}$$

where x_{t-1} is an observable $(k+1)\times 1$ vector of stochastic explanatory variables, such that $x_{t-1} = [1, x_{1,t-1}, ..., x_{k,t-1}]'$, $f(\cdot)$ is a scalar function, $\beta \in \mathbb{R}^{(k+1)}$ is an unknown vector of parameters and

$$\varepsilon_t \mid X \sim F_t(. \mid X)$$

where as before $F_t(. \mid X)$ is a distribution function and $X = [x_0, ..., x_{n-1}]$ is an $n \times (k+1)$ matrix. Once again, we suppose that the error terms process $\{\varepsilon_t, t = 1, 2, \cdots\}$ is a strict conditional mediangale, such that

$$P[\varepsilon_t > 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = P[\varepsilon_t < 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = \frac{1}{2}, \tag{9}$$

with

$$\boldsymbol{\varepsilon}_0 = \{\emptyset\}, \quad \boldsymbol{\varepsilon}_{t-1} = \{\varepsilon_1, \cdots, \varepsilon_{t-1}\}, \quad \text{for} \quad t \geq 2$$

and where (9) entails that $\varepsilon_t \mid X$ has no mass at zero, *i.e.* $P[\varepsilon_t = 0 \mid X]$=0 for all t. We do not require that the parameter vector β be identified. We consider the problem of testing the null hypothesis

$$H(\beta_0) : \beta = \beta_0 \tag{10}$$

against the alternative hypothesis

$$H(\beta_1) : \beta = \beta_1. \tag{11}$$

A test for $H(\beta_0)$ against $H(\beta_1)$ can be constructed as in Section 2.1. First, we note that model (8) is equivalent to the transformed model

$$\tilde{y}_t = g(x_{t-1},\, \beta,\, \beta_0) + \varepsilon_t,$$

where $\tilde{y}_t = y_t - f(x_{t-1}, \beta_0)$ and $g(x_{t-1}, \beta, \beta_0) = f(x_{t-1}, \beta) - f(x_{t-1}, \beta_0)$. Thus, testing $H(\beta_0)$ against $H(\beta_1)$ is equivalent to testing

$$\bar{H}_0 : g(x_{t-1},\, \beta, \beta_0) = 0,\ \text{for } t = 1, ..., n,$$

against

$$\bar{H}_1 : g(x_{t-1}, \beta, \beta_0) = f(x_t, \beta_1) - f(x_t, \beta_0),\ \text{for } t = 1, ..., n.$$

For $\tilde{U}(n) = (s(\tilde{y}_1), ..., s(\tilde{y}_n))'$, where, for $1 \leq t \leq n$,

$$s(\tilde{y}_t) = \begin{cases} 1,\ if\ \tilde{y}_t \geq 0 \\ 0,\ if\ \tilde{y}_t < 0 \end{cases} ,$$

the likelihood function of new random sample $\{s(\tilde{y})_t\}_{t=1}^{n}$ conditional on X is given by:

$$L\left(\tilde{U}(n), \beta\right) = P\left[s(\tilde{y}_1) = \tilde{s}_1, ..., s(\tilde{y}_n) = \tilde{s}_n \mid X\right] = \prod_{t=1}^{n} P\left[s(\tilde{y}_t) = \tilde{s}_t \mid \tilde{\mathrm{S}}_{t-1}, X\right],$$

with

$$\tilde{\mathrm{S}}_0 = \{\emptyset\}\,,\ \tilde{\mathrm{S}}_{t-1} = \{s(\tilde{y}_1) = \tilde{s}_1, ..., s(\tilde{y}_{t-1}) = \tilde{s}_{t-1}\}\,,\ \text{for } t \geq 2,$$

and

$$P\left[s(\tilde{y}_1) = \tilde{s}_1 \mid \tilde{\mathrm{S}}_0, X\right] = P\left[s(\tilde{y}_1) = \tilde{s}_1 \mid X\right],$$

where each $\tilde{s}_t$, for $1 \leq t \leq n$, takes two possible values 0 and 1. Thus, we can use the result of proposition 1 to derive a sign-based test for the null hypothesis $H(\beta_0)$ against the alternative hypothesis $H(\beta_1)$, which leads to the following proposition:

Proposition 2 *Under assumptions (2) and (8),let $H(\beta_0)$ and $H(\beta_1)$ be defined by* (10) - (11),

$$SN_n(\beta_0|\beta_1) = \sum_{t=1}^{n} a_t(\beta_0|\beta_1) s\left(y_t - f(x_{t-1}, \beta_0)\right)$$

where, for $t = 1, ..., n$,

$$a_t(\beta_0|\beta_1) = \ln\left\{\frac{P\left[\tilde{y}_t \geq 0 \mid \tilde{S}_{t-1}, X\right]}{P\left[\tilde{y}_t < 0 \mid \tilde{S}_{t-1}, X\right]}\right\},$$

and suppose the constant $c_1(\beta_0,\, \beta_1)$ satisfies $P\left[\sum\limits_{t=1}^{n} a_t(\beta_0|\beta_1)\ s\left(y_t - f(x_t, \beta_0)\right) > c_1(\beta_0,\, \beta_1)\right] = \alpha$ under $H(\beta_0)$, with $0 < \alpha < 1$. Then the test that rejects $H(\beta_0)$ when

$$SN_n(\beta_0|\beta_1) > c_1(\beta_0,\, \beta_1)$$

is most powerful for testing $H(\beta_0)$ *against* $H(\beta_1)$ *among level-*α *tests based on the signs* $\big(s(\tilde{y}_1), ..., s(\tilde{y}_n)\big)'$.

As in Section 2.1, to make the calculation of the weight function $a_t(\beta_0|\beta_1)$ that depends on the terms $P\left[\tilde{y}_t \geq 0 \mid \underline{\mathrm{S}}_{t-1}, X\right]$ and $P\left[\tilde{y}_t < 0 \mid \underline{\mathrm{S}}_{t-1}, X\right]$ feasible, we consider the following assumption.

Assumption A2: *Let* $\{\tilde{y}_t, t = 0, 1, \cdots\}$ *follow a Markov process of order one. Then under the alternative hypothesis, the sign process* $\{s(\tilde{y}_t)\}_{t=0}^{\infty}$ *is a Markov process of the same order.*

Under Assumption **A2**, the terms $P\left[\tilde{y}_t \geq 0 \mid \tilde{\mathrm{S}}_{t-1}, X\right]$ and $P\left[\tilde{y}_t < 0 \mid \tilde{\mathrm{S}}_{t-1}, X\right]$ simplify and can be expressed as follows:

$$\begin{cases} P\left[\tilde{y}_t \geq 0 \mid \tilde{\mathrm{S}}_{t-1}, X\right] = P\left[\tilde{y}_t \geq 0 \mid \tilde{y}_{t-1} \geq 0, X\right]^{s(\tilde{y}_{t-1})} P\left[\tilde{y}_t \geq 0 \mid \tilde{y}_{t-1} < 0, X\right]^{1-s(\tilde{y}_{t-1})}, \\ \\ P\left[\tilde{y}_t < 0 \mid \tilde{\mathrm{S}}_{t-1}, X\right] = P\left[\tilde{y}_t < 0 \mid \tilde{y}_{t-1} \geq 0, X\right]^{s(\tilde{y}_{t-1})} P\left[\tilde{y}_t < 0 \mid \tilde{y}_{t-1} < 0, X\right]^{1-s(\tilde{y}_{t-1})}. \end{cases}$$

The use of the new expressions of the probabilities $P\left[\tilde{y}_t \geq 0 \mid \tilde{\mathrm{S}}_{t-1}, X\right]$ and $P\left[\tilde{y}_t < 0 \mid \tilde{\mathrm{S}}_{t-1}, X\right]$ simplifies the calculation of the test statistic $SN_n(\beta_0|\beta_1)$. We have the following results:

Corollary 2 *Under assumptions (2) and (1), let* $H(\beta_0)$ *and* $H(\beta_1)$ *be defined by* (10) - (11),

$$\widehat{SN}_n(\beta_0|\beta_1) = \sum_{t=1}^{n} \tilde{a}_t(\beta_0|\beta_1) s(y_t - f(x_{t-1}, \beta_0)) + \sum_{t=1}^{n} \tilde{b}_t(\beta_1)\ s(y_t - f(x_{t-1}, \beta_0)) s(y_{t-1} - f(x_{t-2}, \beta_0)),$$

where

$$\tilde{a}_1(\beta_0|\beta_1) = \ln\left\{\frac{1 - P\left[\varepsilon_1 < f(x_0, \beta_0) - f(x_0, \beta_1) \mid X\right]}{P\left[\varepsilon_1 < f(x_0, \beta_0) - f(x_0, \beta_1) \mid X\right]}\right\},\ \tilde{b}_1(\beta_0|\beta_1) = 0$$

and for $t = 2, ..., n$,

$$\tilde{a}_t(\beta_0|\beta_1) = \ln\left\{\frac{1 - \frac{P[\varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1),\ \varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X]}{P[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X]}}{\frac{P[\varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1),\ \varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X]}{P[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X]}}\right\}$$

$$\tilde{b}_t(\beta_0|\beta_1) = \ln\left\{\frac{1 - \left(\frac{P\left[\varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1) | X\right]}{1 - P\left[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]} - \frac{P\left[\ \varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1),\ \varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1) | X\right]}{1 - P\left[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]}\right)}{\frac{P\left[\varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1) | X\right]}{1 - P\left[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]} - \frac{P\left[\ \varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1),\ \varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1) | X\right]}{1 - P\left[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]}}\right\}$$

$$- \ln\left\{\frac{1 - \frac{P\left[\varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1),\ \varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]}{P\left[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]}}{\frac{P\left[\varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1),\ \varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]}{P\left[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]}}\right\}$$

and suppose the constant $\tilde{c}_1(\beta_0, \beta_1)$ *satisfies* $P\left[\widehat{SN}_n(\beta_0|\beta_1) > \tilde{c}_1(\beta_0, \beta_1)\right] = \alpha$ *under* $H(\beta_0)$, *with* $0 < \alpha <$

1. *Then the test that rejects* $H(\beta_0)$ *when*

$$\widehat{SN}_n(\beta_0|\beta_1) > \tilde{c}_1(\beta_0, \beta_1)$$

is most powerful for testing $H(\beta_0)$ *against* $H(\beta_1)$ *among level-*α *tests based on the signs* $\big(s(\tilde{y}_1), ..., s(\tilde{y}_n)\big)'$.

If we consider a linear function $f(x'_{t-1}, \beta) = \beta' x_{t-1}$, as before we may suppose that ε_t for $t = 1, \cdots, n$ follow $N(0, 1)$, which allows us to evaluate the bivariate probabilities by utilizing the "jointly symmetric" copula or the multivariate Student's t distribution with the identity matrix imposed. Then the test statistic for the null hypothesis $H(\beta_0)$ against the alternative $H(\beta_1)$ is given by

$$\widehat{SN}_n(\beta_0|\beta_1) = \sum_{t=1}^{n} \tilde{a}_t(\beta_0|\beta_1) s(y_t - \beta'_0 x_{t-1}) + \sum_{t=1}^{n} \tilde{b}_t(\beta_1) s(y_t - \beta'_0 x_{t-1}) s(y_{t-1} - \beta'_0 x_{t-2}),$$

where

$$\tilde{a}_1(\beta_0|\beta_1) = \ln\left\{ \frac{\Phi\left((\beta_1 - \beta_0)' x_0\right)}{1 - \Phi\left((\beta_1 - \beta_0)' x_0\right)} \right\}, \quad \tilde{b}_1(\beta_0|\beta_1) = 0,$$

and for $t = 2, ..., n$,

$$\tilde{a}_t(\beta_0|\beta_1) = \ln\left\{ \frac{1 - \frac{C^{JS}(\Phi((\beta_0-\beta_1)' x_{t-1}), \Phi((\beta_0-\beta_1)' x_{t-2}))}{1-\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)}}{\frac{C^{JS}(\Phi((\beta_0-\beta_1)' x_{t-1}), \Phi((\beta_0-\beta_1)' x_{t-2}))}{1-\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)}} \right\},$$

$$\tilde{b}_t(\beta_0|\beta_1) = \ln\left\{ \frac{1 - \left(\frac{1-\Phi\left((\beta_1-\beta_0)' x_{t-1}\right)}{\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)} - \frac{C^{JS}(\Phi((\beta_0-\beta_1)' x_{t-1}), \Phi((\beta_0-\beta_1)' x_{t-2}))}{\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)} \right)}{\frac{1-\Phi\left((\beta_1-\beta_0)' x_{t-1}\right)}{\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)} - \frac{C^{JS}(\Phi((\beta_0-\beta_1)' x_{t-1}), \Phi((\beta_0-\beta_1)' x_{t-2}))}{\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)}} \right\}$$

$$- \ln\left\{ \frac{1 - \frac{C^{JS}(\Phi((\beta_0-\beta_1)' x_{t-1}), \Phi((\beta_0-\beta_1)' x_{t-2}))}{1-\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)}}{\frac{C^{JS}(\Phi((\beta_0-\beta_1)' x_{t-1}), \Phi((\beta_0-\beta_1)' x_{t-2}))}{1-\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)}}, \right\}$$

where $\Phi(.)$ is the standard normal distribution function and $C^{JS}(u_1, u_2)$ is the "jointly-symmetric" bivariate copula with uniformly distributed marginals. As in Section 2.1, the test statistic $\widehat{SN}_n(\beta_0|\beta_1)$ depends on a particular alternative hypothesis β_1. In practice, the latter is unknown, which makes the proposed POS test infeasible. However, in Section 3 we will suggest an adaptive approach [see Dufour and Taamouti (2010)] which can be used to choose an optimal alternative hypothesis at which the power of the test is maximized.

3 Choice of the optimal alternative hypothesis

Point-optimal tests depend on the alternative $\beta = \beta_1$, which in practice is unknown. Formally, the test statistic $\widehat{SN}_n(\beta_0|\beta_1)$ for testing the linear full-coefficient hypothesis (10) is a function of β_1

$$\widehat{SN}_n(\beta_0|\beta_1) = \sum_{t=1}^{n} \tilde{a}_t(\beta_0|\beta_1)s(y_t - x'_{t-1}\beta_0) + \sum_{t=1}^{n} \tilde{b}_t(\beta_1)s(y_t - x'_{t-1}\beta_0)s(y_{t-1} - x'_{t-2}\beta_0),$$

which in turn implies that its power function, say $\Pi(\beta_0, \beta_1)$, is also a function of β_1. Therefore, the choice of the alternative β_1 has a direct impact on its power function. In other words,

$$\Pi(\beta_0, \beta_1) = P[\widehat{SN}_n(\beta_0|\beta_1) > \tilde{c}_1(\beta_0, \beta_1) \mid H(\beta_1)],$$

where $\tilde{c}_1(\beta_0, \beta_1)$ satisfies the constraint

$$P[\widehat{SN}_n(\beta_0|\beta_1) > \tilde{c}_1(\beta_0, \beta_1)|H(\beta_0)] \leq \alpha.$$

Our objective is to choose the value of β_1 at which the power of the POS-based test statistic is maximized and is close to that of the power envelope. This can be accomplished in a number of ways. Dufour and Taamouti (2010) suggest an adaptive approach based on the split-sample technique [see Dufour and Jasiak (2001)] for estimating the optimal alternative to make size control easier and maximize the power. For a review of adaptive approach for parametric tests with non-standard distributions see Dufour and Taamouti (2003) and Dufour et al. (2008).

This approach consists of splitting the sample into two independent parts, where the alternative β_1 is estimated using the first part, while the POS test-statistic $\widehat{SN}_n(\beta_0|\beta_1)$ is calculated using the second part of the sample, along with the alternative β_1 estimated using the first part. By adopting this technique, size control is easier and the power function of the POS-test traces out the power envelope. Let $n = n_1 + n_2$, $y = (y'_{(1)}, y'_{(2)})'$, $X = (X'_{(1)}, X'_{(2)})'$, and $\varepsilon = (\varepsilon'_{(1)}, \varepsilon'_{(2)})'$, where $y_{(i)}$, $X_{(i)}$ and $u_{(i)}$ for $i \in \{1, 2\}$ each have n_i rows. The first n_1 observations of y and X can thus be denoted by $y_{(1)}$ and $X_{(1)}$, which are used for estimating the alternative hypothesis β_1 with the OLS estimator:

$$\hat{\beta}_{(1)} = (X'_{(1)}X_{(1)})^{-1}X'_{(1)}y_{(1)}.$$

Alternatively, in the case of extreme outliers other robust estimators that are less sensitive to outliers can be utilized [see Maronna et al. (2019) for a review of robust estimators]. Since $\hat{\beta}_{(1)}$ is independent of $X_{(2)}$, the last n_2 observations can be used to calculate the test statistic and obtain a valid POS test

$$\widehat{SN}_n(\beta_0|\beta_{(1)}) = \sum_{t=n_1+1}^{n} \tilde{a}_t(\beta_0|\beta_{(1)})s(y_t - x'_{t-1}\beta_0) + \sum_{t=n_1+1}^{n} \tilde{b}_t(\beta_{(1)})s(y_t - x'_{t-1}\beta_0)s(y_{t-1} - x'_{t-2}\beta_0),$$

Different choices for n_1 and n_2 is possible. However, as Dufour and Taamouti (2010) have noted, the number of observations retained for the first and the second sub-samples has a direct impact on the power of the test, and a more powerful test is obtained when a relatively small number of observations is used for estimating the alternative and more observations are saved for calculating the test statistic. Having conducted a simulation study to compare the power-curves of split-sample-based POS tests to the power envelope, they find that using approximately 10% of the sample to estimate the alternative yields a power which is very close to that of the power envelope. Therefore, we follow Dufour and Taamouti (2010) by using the first 10% of the sample to estimate the alternative and the remaining 90% to calculate the test statistic.

4 POS confidence regions

In this Section, we follow Dufour and Taamouti (2010) and Coudin and Dufour (2009) to discuss the process of building confidence regions at a given significance level α, say $C_\beta(\alpha)$, for a vector (sub-vector) of the unknown parameters β using the proposed POS test. We consider again the linear regression (8) and suppose we wish to test the null hypothesis (10) against the alternative hypothesis (11). Formally, the idea involves finding all the values of $\beta_0 \in \mathbb{R}^{(k+1)}$ such that

$$\widehat{SN}_n(\beta_0|\,\beta_1) = \sum_{t=1}^{n} \tilde{a}_t(\beta_0 \mid \beta_1) s(y_t - \beta_0' x_{t-1}) + \sum_{t=1}^{n} \tilde{b}_t(\beta_1) s(y_t - \beta_0' x_{t-1}) s(y_{t-1} - \beta_0' x_{t-2}) < \tilde{c}_1(\beta_0, \beta_1) \quad (12)$$

where the critical value $\tilde{c}_1(\beta_0, \beta_1)$ satisfies the constraint

$$P\left[\widehat{SN}_n(\beta_0|\,\beta_1) > \tilde{c}_1(\beta_0, \beta_1) | \beta = \beta_0\right] \leq \alpha$$

Thus, the confidence region $C_\beta(\alpha)$ of the vector of parameters β is defined as

$$C_\beta(\alpha) = \left\{\beta_0 : \widehat{SN}_n(\beta_0|\,\beta_1) < \tilde{c}_1(\beta_0, \beta_1) | P[\widehat{SN}_n(\beta_0|\,\beta_1) > \tilde{c}_1(\beta_0, \beta_1) | \beta = \beta_0] \leq \alpha\right\}.$$

Given the confidence region $C_\beta(\alpha)$, confidence intervals for the components of vector β can be obtained using the projection techniques. Confidence sets in the form of transformations T of $\beta \in \mathbb{R}^m$, $T(C_\beta(\alpha))$ for $m \leq (k+1)$ can easily be found using the said techniques. Since, for any set $C_\beta(\alpha)$

$$\beta \in C_\beta(\alpha) \implies T(\beta) \in T(C_\beta(\alpha)), \quad (13)$$

we have

$$P[\beta \in C_\beta(\alpha)] \geq 1 - \alpha \implies P[T(\beta) \in T(C_\beta(\alpha))] \geq 1 - \alpha, \quad (14)$$

where

$$T(C_\beta(\alpha)) = \{\delta \in \mathbb{R}^m : \exists \beta \in C_\beta(\alpha), T(\beta) = \delta\}.$$

From (13) and (14), it is evident that the set $T(C_\beta(\alpha))$ is a conservative confidence set for $T(\beta)$ with level $1-\alpha$. If $T(\beta)$ is a scalar, then we have

$$P[\inf\{T(\beta_0), \text{ for } \beta_0 \in C_\beta(\alpha)\} \leq T(\beta) \leq \sup\{T(\beta_0), \text{ for } \beta_0 \in C_\beta(\alpha)\}] > 1-\alpha.$$

To obtain valid conservative confidence intervals for the individual component β_j in (8) under assumption (2), we follow Coudin and Dufour (2009) by implementing a global numerical optimization search algorithm to solve the problem

$$\min_{\beta \in \mathbb{R}^{(k+1)}} \beta_j \;\; s.c. \;\; \widehat{SN}_n(\beta_0 \mid \beta_1) < \tilde{c}_1(\beta_0, \beta_1), \qquad \max_{\beta \in \mathbb{R}^{(k+1)}} \beta_j \;\; s.c. \;\; \widehat{SN}_n(\beta_0 \mid \beta_1) < \tilde{c}_1(\beta_0, \beta_1) \tag{15}$$

where the critical value $c(\beta_0, \beta_1)$ at level α, is computed using B replications of the statistic $\widehat{SN}_n^{(i)}(\beta_0 \mid \beta_1)$ under the null hypothesis and finding its $(1-\alpha)$ quantile. Using the projection technique, multiple tests maintain control of the overall level when performed on an arbitrary number of hypotheses.

4.1 Numerical illustration

Following Coudin and Dufour (2009), we illustrate the projection technique by generating a process with sample size $n = 500$, such that

$$y_t = \beta_0 + \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1} + \varepsilon_t \overset{i.i.d}{\sim} \begin{cases} N(0,1) & \text{with probability } 0.95 \\ N(0,100^2) & \text{with probability } 0.05 \end{cases}$$

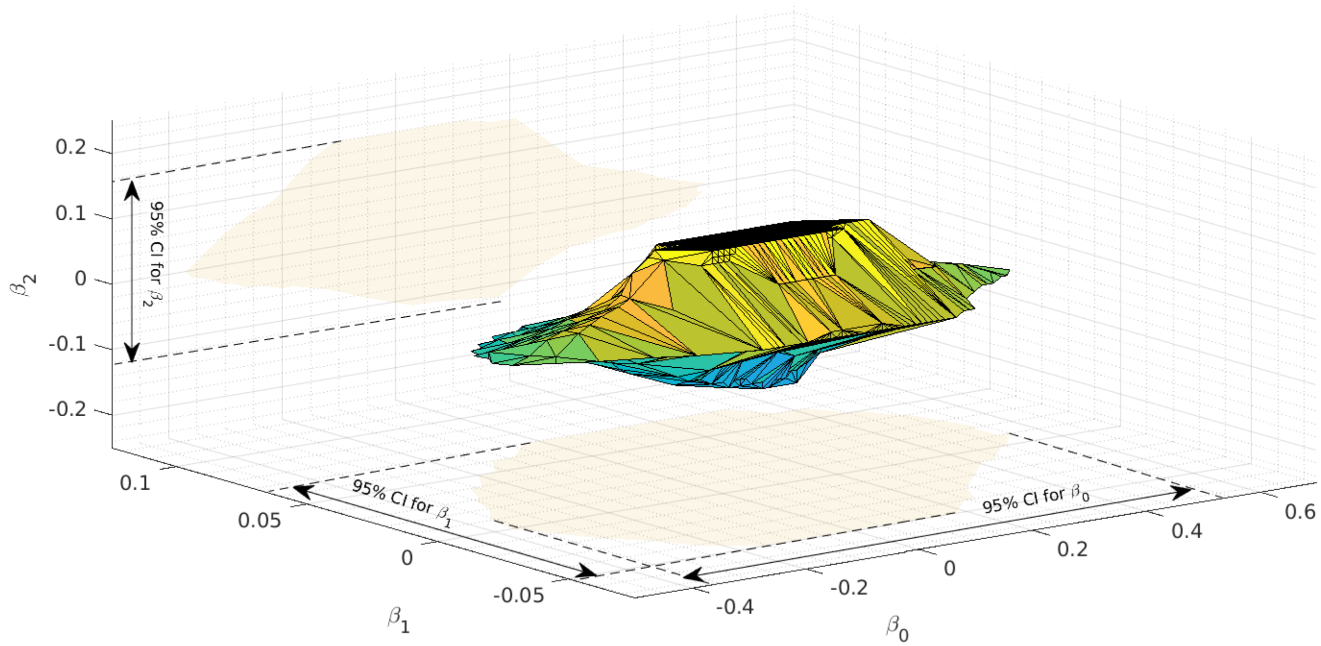
where $\beta_0 = \beta_1 = \beta_2 = 0$ and

$$x_{1,t} = \theta_1 x_{1,t-1} + u_{1,t}$$
$$x_{2,t} = \theta_2 x_{2,t-1} + u_{2,t}$$

with $\theta_1 = \theta_2 = 0.9$ The initial values of x_1 and x_2 are respectively given by: $x_{1,0} = \frac{u_{1,0}}{\sqrt{1-\theta_1^2}}$ and $x_{2,0} = \frac{u_{2,0}}{\sqrt{1-\theta_2^2}}$, and $u_{1,t}$ and $u_{2,t}$ are generated from $N(0,1)$.

The exact inference procedure is conducted with $B = 1,000$ replications of the test statistic under the null hypothesis. As β is a vector in three-dimensional space, the confidence region and the projections can be illustrated graphically. The tests of $H_0(\beta^*) : \beta = \beta^*$ are performed on a grid for $\beta^* = (\beta_0^*, \beta_1^*, \beta_2^*)$. Due to the curse of dimensionality encountered in the process of creating a grid for the parameters, the *simulated annealing optimization algorithm* is initially used to solve problem (15) for each parameter β_i, to obtain a realistic size dimension of the grid [see Goffe et al. (1994) for a review of the simulated annealing algorithm].

Figure 1: 95% confidence region for the unknown vector $\beta = (\beta_0, \beta_1, \beta_2)$ obtained by searching a three-dimensional grid β^* using the 10% SS-POS test.



Note: The shaded regions on the $\beta_0 - \beta_1$ and $\beta_2 - \beta_1$ planes are the shadows casted by the three-dimensional confidence region, which simplify the visual identification of the 95% confidence intervals for each parameter β_i.

The optimizations were performed using MATLAB software on a high-performance computing (HPC) cluster, by utilizing six nodes each equipped with Intel(R) Xeon(R) 16-core processors (2.40GHz). The simulated annealing algorithm's speed of adjustment was set to 0.25, with a temperature reduction factor of 75%, an initial temperature of 50 and a convergence criteria of 0.01. All algorithms converged in less than an hour. Once the global maximum and minimum for each parameter β_i were obtained, the grid was constructed by the Cartesian product of the linearly spaced distance between the β_i's maxima and minima.

Table 1: Comparison of the 95% confidence intervals obtained for the unknown parameters β_0, β_1 and β_2 using the 10% SS-POS-test, with those achieved using the T-test and T-test based on White (1980) variance correction.

		OLS	White	10% SS-POS
β_0	95% CI	**[-0.01, -0.00]**	$[-0.01, 0.00]$	$[-0.37, 0.55]$
β_1	95% CI	**[-1.04, -0.60]**	**[-1.09, -0.56]**	$[-0.05, 0.07]$
β_2	95% CI	**[0.47, 0.67]**	**[0.45, 0.69]**	$[-0.12, 0.16]$

Note: The confidence intervals in bold do not contain the value of zero and imply significance at the 5% level.

It is evident that the 10% split-sample POS-based test outperforms the T-test and the T-test based on white (1980) variance correction test, as the former correctly fails to reject the null hypothesis of orthogonality at the 5% level, whereas the latter two tests reject the null hypothesis in favor of the alternative for almost all parameters.

5 Monte Carlo study

In this Section, we provide simulation results that illustrate the performance of the POS-based tests proposed in the previous Sections. We have limited our results to two groups of data generating processes (DGPs) which correspond to different symmetric and asymmetric distributions and different forms of heteroskedasticity.

5.1 Simulation setup

We assess the performance of the proposed 10% SS-POS tests in terms of size and power, by considering various DGPs with symmetric and asymmetric distributions and different forms of heteroskedasticity. The DGPs under consideration are supposed to mimic different scenarios that are often encountered in practical settings within the domains of predictive regressions. The performance of the 10% SS-POS tests is compared to that of a few other tests, by considering the following linear predictive regression model

$$y_t = \beta x_{t-1} + \varepsilon_t \tag{16}$$

where β is an unknown parameter. Furthermore, we follow Mankiw and Shapiro (1986) by assuming that x_t is a stationary AR(1) process

$$x_t = \theta x_{t-1} + u_t \tag{17}$$

such that u_t are mutually independent, and each u_t is independent of x_{t-k} for $k \geq 1$. Moreover, the disturbances (ε_t, u_t) are distributed as bivariate normal, with the contemporaneous covariance matrix

$$\Sigma_{\varepsilon u} = \begin{bmatrix} 1 & \sigma_{\varepsilon u} \\ \sigma_{\varepsilon u} & \sigma_u^2 \end{bmatrix}$$

Therefore, there is feedback from u_t to x_t through ε_t, which implies that $corr(\varepsilon_t, x_{t+k}) \neq 0$ for $k \geq 0$. Thus, as the disturbance vector $[\varepsilon_1, \cdots, \varepsilon_n]'$ is not independent of the regressor vector $[x_0, \cdots, x_{n-1}]'$, the OLS estimator is biased in finite-samples and the T-statistic has a non-standard distribution. Mankiw and Shapiro (1986) perform an extensive simulations exercise by considering different values of the correlation between u_t and ε_t (say ρ) and find that in small samples, as θ and ρ approach unity, the T-test using asymptotic critical values leads to oversized tests; however, this size distortion improves as $n \to \infty$.

To compare the performance of certain parametric and non-parametric tests to that of the POS-based tests, the data is generated from model (16), with the stationary process x_t specified as (17) and by further setting

$$u_t = \rho\varepsilon_t + w_t\sqrt{1-\rho^2} \tag{18}$$

for $\rho = 0$, 0.1, 0.5, 0.9, where ε_t and w_t are assumed to be independent. The initial value of x is given by: $x_0 = \frac{w_0}{\sqrt{1-\theta^2}}$. Further, w_t are generated from $N(0,1)$ and we assign $\theta = 0.9$.

The errors ε_t are i.n.i.d and are categorized by two groups in our simulation study. In the first group, we consider DGPs where the residuals ε_t possess symmetric and asymmetric distributions:

1. normal distribution: $\varepsilon_t \sim N(0,1)$;

2. Cauchy distribution: $\varepsilon_t \sim Cauchy$;

3. Student t distribution with two degrees of freedom: $\varepsilon_t \sim t(2)$;

4. Mixture of normal and Cauchy distributions: $\varepsilon_t \sim s_t \mid \varepsilon_t^C \mid -(1-s_t) \mid \varepsilon_t^N \mid$, where ε_t^C follows Cauchy distribution, ε_t^N follows $N(0,1)$ distribution, and

$$P(s_t = 1) = P(s_t = 0) = \frac{1}{2}.$$

The second group of DGPs represents different forms of heteroskedasticity:

5. break in variance:

$$\varepsilon_t \sim \begin{cases} N(0,1) & \text{for } t \neq 25 \\ \sqrt{1000}N(0,1) & \text{for } t = 25 \end{cases};$$

6. exponential variance: $\varepsilon_t \sim N(0, \sigma_\varepsilon^2(t))$ and $\sigma_\varepsilon(t) = \exp(0.5t)$;

7. GARCH(1, 1) plus jump variance:

$$\sigma_\varepsilon^2(t) = 0.00037 + 0.0888\varepsilon_{t-1}^2 + 0.9024\sigma_\varepsilon^2(t-1)\,,$$

$$\varepsilon_t \sim \begin{cases} N(0, \sigma_\varepsilon^2(t)) & \text{for } t \neq 25 \\ 50N(0, \sigma_\varepsilon^2(t)) & \text{for } t = 25 \end{cases} ;$$

8. nonstationary GARCH(1, 1) variance: $\varepsilon_t \sim N(0, \sigma_\varepsilon^2(t))$ and

$$\sigma_\varepsilon^2(t) = 0.75\varepsilon_{t-1}^2 + 0.75\sigma_\varepsilon^2(t-1)\,.$$

We implement the POS-based test and other tests, which are intended to be robust against heteroskedasticity and non-normality, to test the null hypothesis of orthogonality - i.e. $H_0 : \beta = 0$. As in Dufour and Taamouti (2010), Monte Carlo simulations are used to compare the size and power of the 10% split-sample POS-based tests (10% SS-POS test hereafter) to those of T-test, T-test based on White (1980) variance correction (hereafter WT-test), and sign-based test proposed by Campbell and Dufour (1995) (CD(1995) test hereafter). The simulation study involves $M_1 = 10,000$ iterations for evaluating the probability distribution of POS test statistic and $M_2 = 5,000$ iterations to estimate the power functions of POS test and other tests. We consider a sample size of $n = 50$ for conducting the simulation exercise. Note that the sign-based test statistic of Campbell and Dufour (1995) possesses a discrete distribution, as a result of which it is not possible (without randomization) to attain test whose size is exactly 5%. In our simulations study, the size of the aforementioned test is 5.95% for $n = 50$.

As in Mankiw and Shapiro (1986), it is further possible to consider values of ρ and θ closer to unity at which the size distortions of T-type tests are magnified. For instance, the size of the T-test in their study is shown to be severely distorted with values of $\theta = 0.999$ and $\rho = 1.0$, given a sample size of $n = 50$. The simulations for the latter scenario can be found in the Appendix for standard normal disturbances. It must be noted that as the exact finite-sample distribution of the POS-based tests are simulated, our tests control size regardless of the values of ρ and θ - the results in figure (8) confirm these findings. It is further evident that although the size distortions for the T-test and T-test based on White (1980) variance correction improve in large samples, these tests still reject the null hypothesis at twice and thrice their nominal level respectively given a sample of $n = 500$ observations.

The DGPs considered in this chapter have been inspired by the simulation exercises conducted in the previous studies [see Mankiw and Shapiro (1986), Campbell and Dufour (1995), Coudin and Dufour (2009) and Dufour and Taamouti (2010)]. The first three DGPs all possess symmetrical distributions that are independent and identical across different observations $t = 1, \cdots, n$. Evidently, as depicted in figure (**??**), the Cauchy and Student's t distribution possess heavier tails in comparison to that of the normal distribution. The standard error of the coefficients are inflated in the presence of heavy tails, as a result of which the power of the T-type tests tend to be poor in comparison to other measures of central tendency (such as the median). Furthermore, the length of the confidence intervals are extended when the data is sampled from heavy tailed distributions. DGP 4 is a mixture of Cauchy and Gaussian distribution; as such, while the errors are independent, they are not identically distributed across different observations [see figure **??**]. DGP 4 is inspired by Magdalinos and Phillips (2009), who note that when x_t is moderately explosive (with $\theta > 1$), the least squares estimator is mixed normal with Cauchy-type tail behavior with an explosive convergence rate. The second group of DGPs covers different forms of heteroskedasticity, such as conditional heteroskedasticity (e.g. stationary and non-stationary GARCH models) and other forms of non-linear dependencies. Dufour and Taamouti (2010) show that under certain forms of heteroskedasticity, T-type tests are not valid; hence, these DGPs fit well within the domains of our study.

5.2 Simulation results

Monte Carlo simulation results are presented in Figures 2-6. These results correspond to different DGPs described in Section 5.1. The figures compare the power of the 10% SS-POS test to the T-test, WT-test, and CD (1995) test. The results are detailed below.

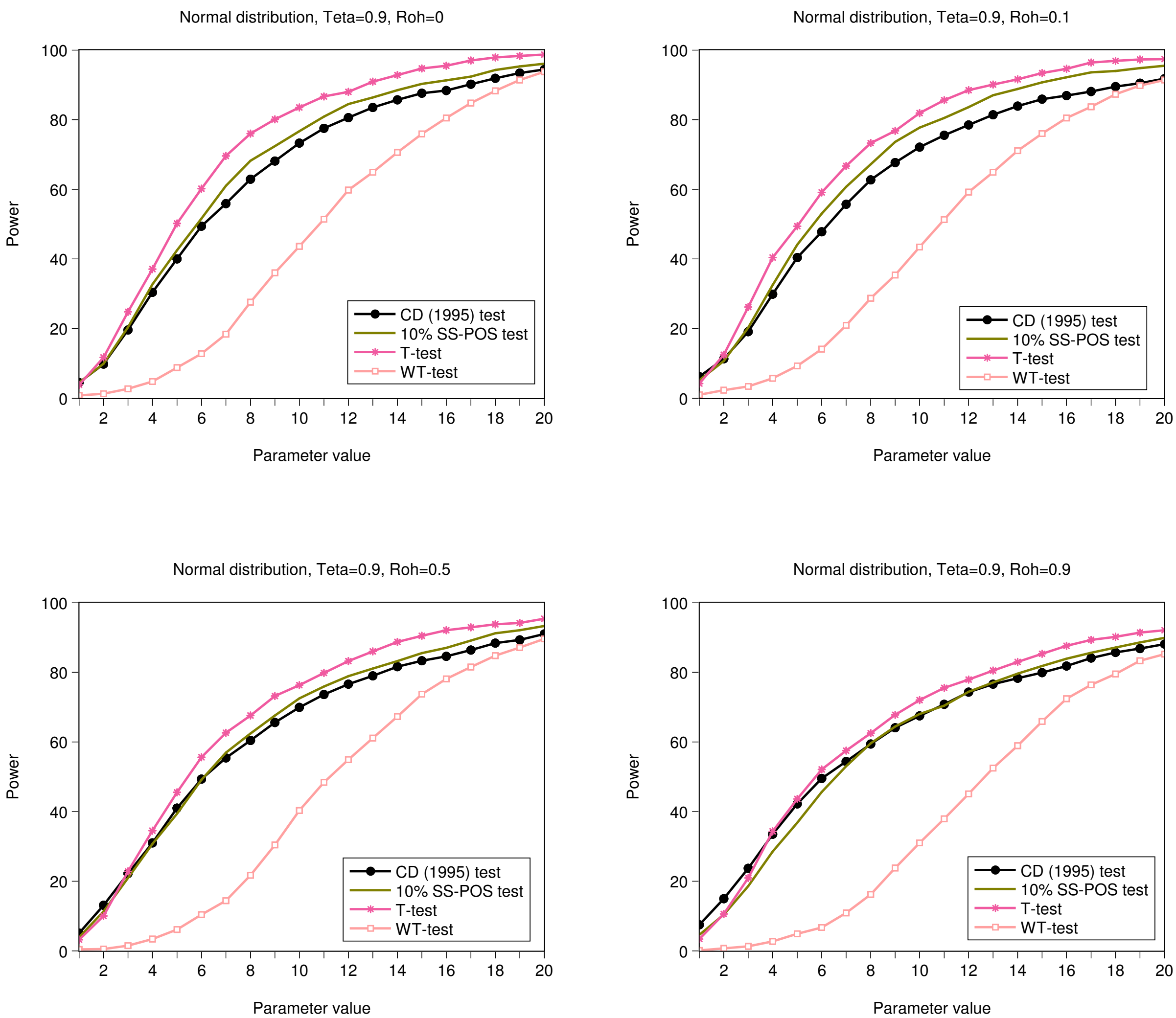
First, Figure 2 compares the power function of the above tests in the case where the error term ε_t in the model (16) is normally distributed. From this we see that all these tests control size, except WT-test which is undersized. We also find that T-test is more powerful than 10% SS-POS test, CD (1995) test, and WT-test. This result is expected since under normality T-test is the most powerful test. However, the power of 10% SS-POS test has the second best power among the other tests. These results are still the same when we increase the correlation coefficient ρ, except that when there high correlation between the error terms ε_t and w_t the power curves of T-test, 10% SS-POS test and CD (1995) test become closer to each other.

Second, Figure 3 corresponds to the cases where the error term ε_t follows Cauchy distribution. From this we see that 10% SS-POS test is more powerful than CD (1995) test, WT-test, and the T-test. It seems that the latter two tests are undersized. 10% SS-POS test and CD (1995) test have much more power than WT-test and T-test for small values (0 and 0.1) of correlation coefficient ρ, but the difference in power decreases when we increase ρ even if it still quite important.

Third, Figure 4 corresponds to the cases where the error term ε_t follows a mixture of normal and Cauchy distributions. The results show that 10% SS-POS test is again more powerful than CD (1995), T-test, and the WT-test. The difference in power is much more significant when the correlation coefficient ρ is smaller.

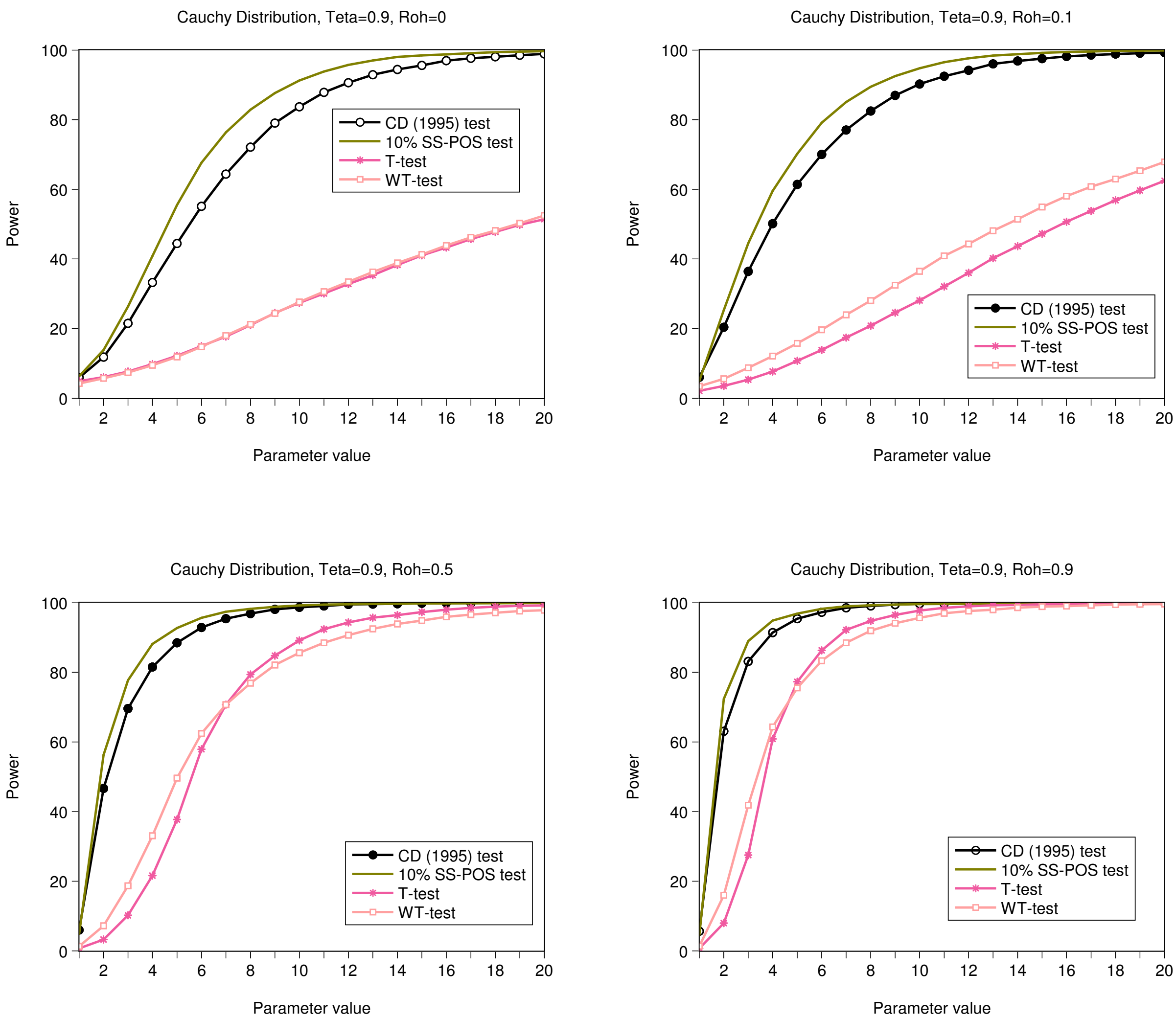
Finally, Figures 5 and 6 compare the power function of the 10% SS-POS test, CD (1995) test, WT-test, and T-test in the case where ε_t follows normal distribution with a break in variance and an exponential variance, respectively. Figure 5 shows that in the presence of break in variance, WT-test and T-test are undesized, whereas 10% SS-POS test and CD (1995) test control size. In addition, 10% SS-POS test has more power that the other tests. The CD (1995) test has the second best power followed by WT-test

Figure 2: Power comparisons: different tests. Normal error distributions with different values of ρ in (18) and $\theta = 0.9$ in (17).



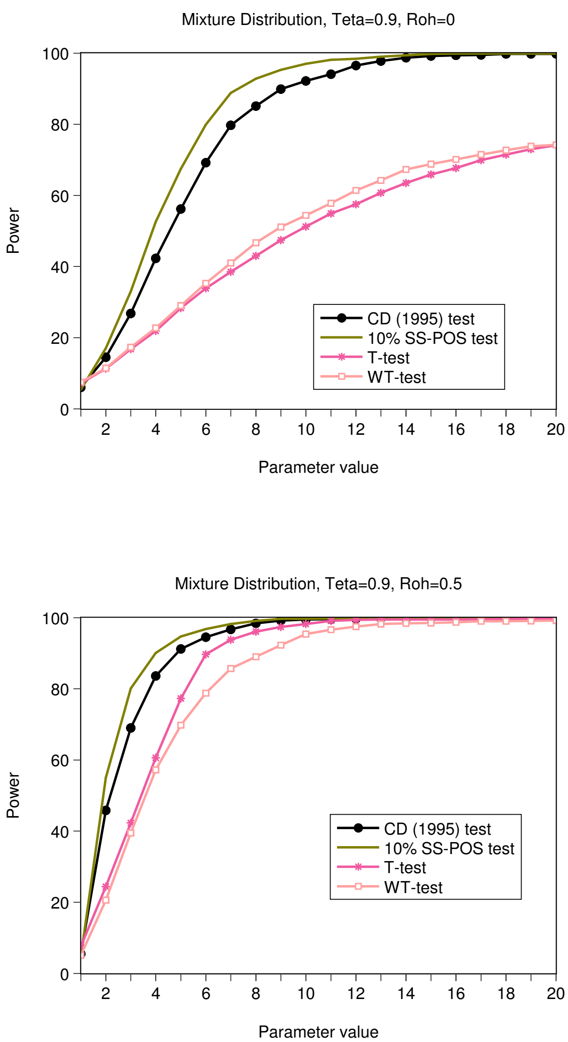
Note: These figures compare the power function of the 10% SS-POS test with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

Figure 3: Power comparisons: different tests. Cauchy error distributions with different values of ρ in (18) and $\theta = 0.9$ in (17).



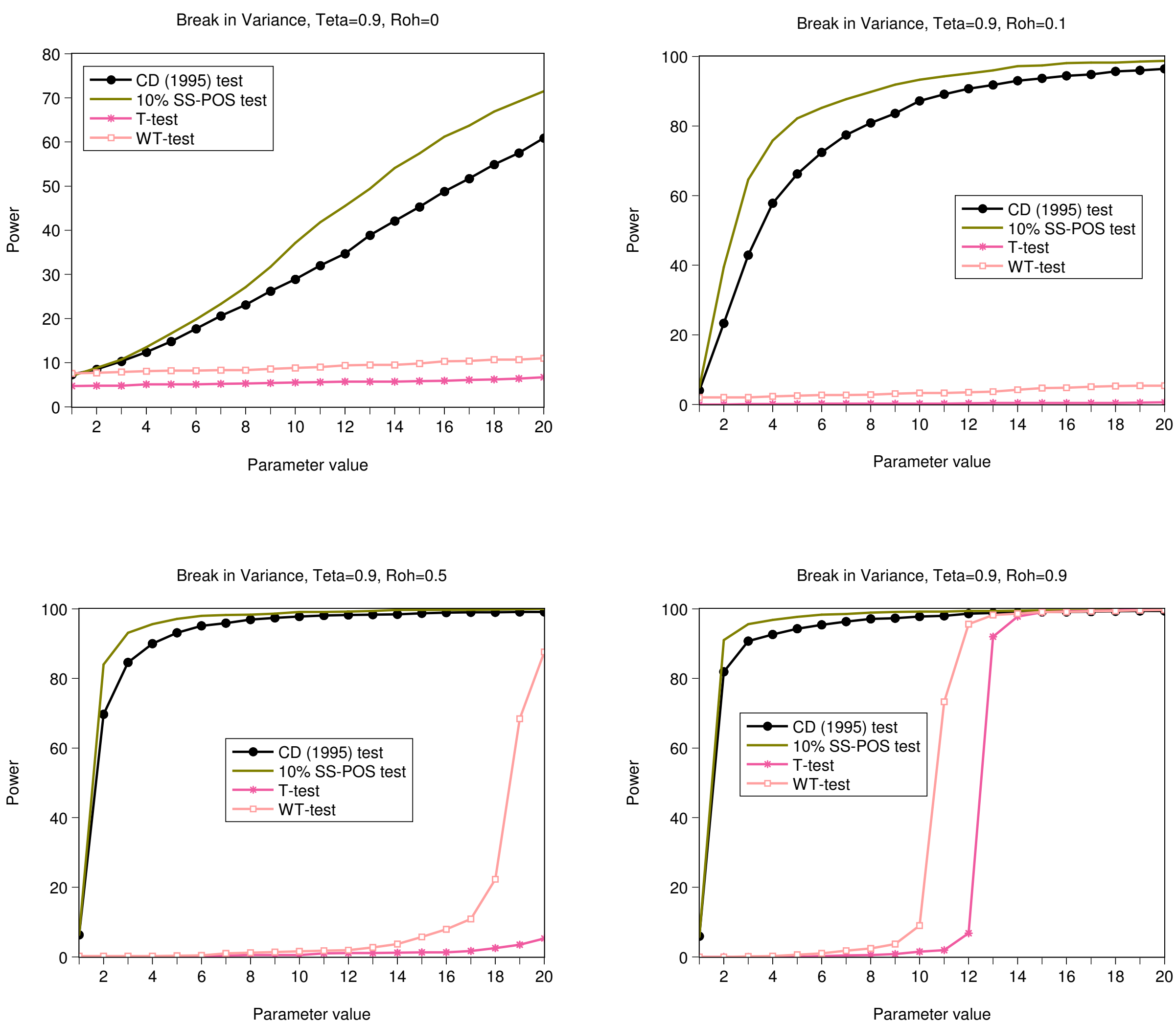
Note: These figures compare the power function of the 10% SS-POS test with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

Figure 4: Power comparisons: different tests. Mixture error distributions with different values of ρ in (18) and $\theta = 0.9$ in (17).



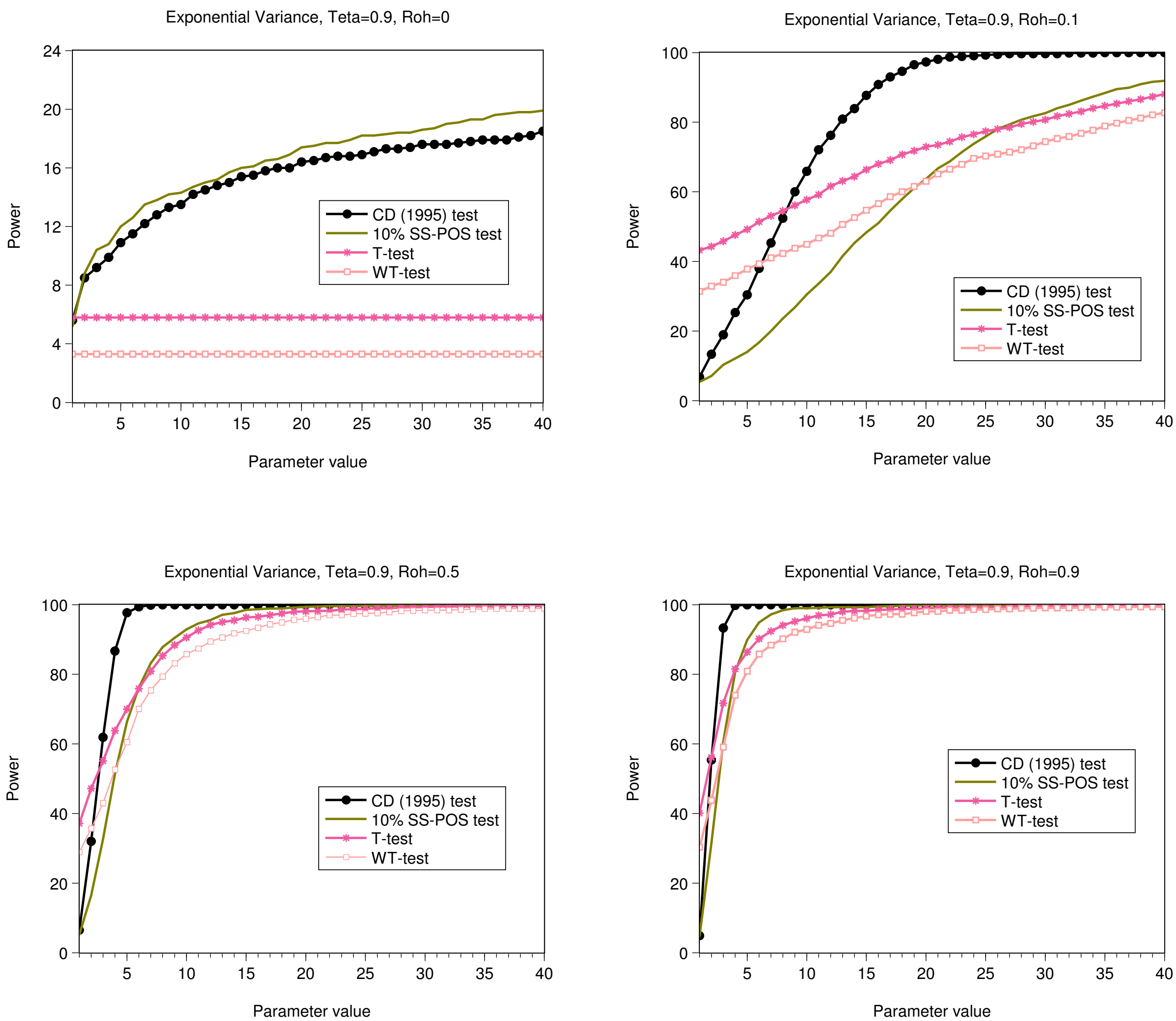
Note: These figures compare the power function of the 10% SS-POS test with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

Figure 5: Power comparisons: different tests. Normal error distributions with break in variance, different values of ρ in (18) and $\theta = 0.9$ in (17).



Note: These figures compare the power function of the 10% SS-POS test with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

Figure 6: Power comparisons: different tests. Normal error distributions with Exp(t) variance, different values of ρ in (18) and $\theta = 0.9$ in (17).



Note: These figures compare the power function of the 10% SS-POS test with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

and T-test. The power of these tests improve when we increase the correlation coefficient ρ. Figure 6 shows that in the case of exponential variance, the WT-test, and T-test are oversized. We find that 10% SS-POS test has more power than CD (1995) test when ρ is equal to zero. However, CD (1995) test becomes more powerful than 10% SS-POS test when correlation coefficient ρ increases. The difference in power between the latter two tests becomes small for higher values of ρ.

6 Empirical application

In this Section, we consider an empirical application of the proposed 10% SS-POS tests to illustrate its practical relevance. Valuation ratios are widely considered as predictors of stock returns and are generally known to be persistent. Therefore, they fit well within the framework of our study. In what follows, we specifically divert our attention to an application in the context of stock return predictability using the said ratios.

6.1 Stock return predictability using valuation ratios

Many studies have investigated the predictive power of valuation ratios on excess stock returns. Dividend-price and earnings-price ratios are among few that were the focus of study in the early 1980s. The attention to these ratios was heightened when Rozeff (1984), Fama and French (1988), and Campbell and Shiller (1988) showed the ratios positive correlation with ex-post stock returns. Fama and French (1988) find that in short horizons dividend yields only explain a small fraction of the variation in time-varying returns, yet in longer horizons (beyond one year) this proportion is significantly increased. Campbell and Shiller (1988) employ a two-variable system approach with the lagged log of the dividend-price ratio together with the lagged real dividend growth rate, to show significant predictive power on stock returns.

These studies are typically performed by regressing the excess returns on a constant and a lagged variable. The conventional T-test is then used to make inference concerning predictability. However, most of these studies are based on the presumption of the stationarity of the predictors, where the T-statistic is approximately normally distributed in large samples. Unfortunately, this is not the case in the presence of highly persistent variables. Even when the predictors are stationary, asymptotic critical values are not a good approximation for those obtained in finite-sample distributions. In the presence of highly persistent predictors, the innovations are greatly correlated with the returns, and thus, the T-statistic has a non-standard distribution which leads to the over-rejection of the null hypothesis of orthogonality [see. Elliott and Stock (1994), Mankiw and Shapiro (1986), Stambaugh (1999) and Campbell and Yogo (2006)].

Most studies address the issue of persistency by making inference based on more accurate approximations of the finite-sample distribution of the test-statistic. This is accomplished either by relying on exact finite-sample theory under the assumption of normality [see. Evans and Savin (1981, 1984) and Stambaugh (1999)] or local-to-unity asymptotics [see Elliott and Stock (1994), Campbell and Yogo (2006) and Torous et al. (2004)]. More recently Taamouti et al. (2014) confirm the predictability power of the valuation ratios using monthly data, in a nonparametric and model-free copula-based Granger causality framework.

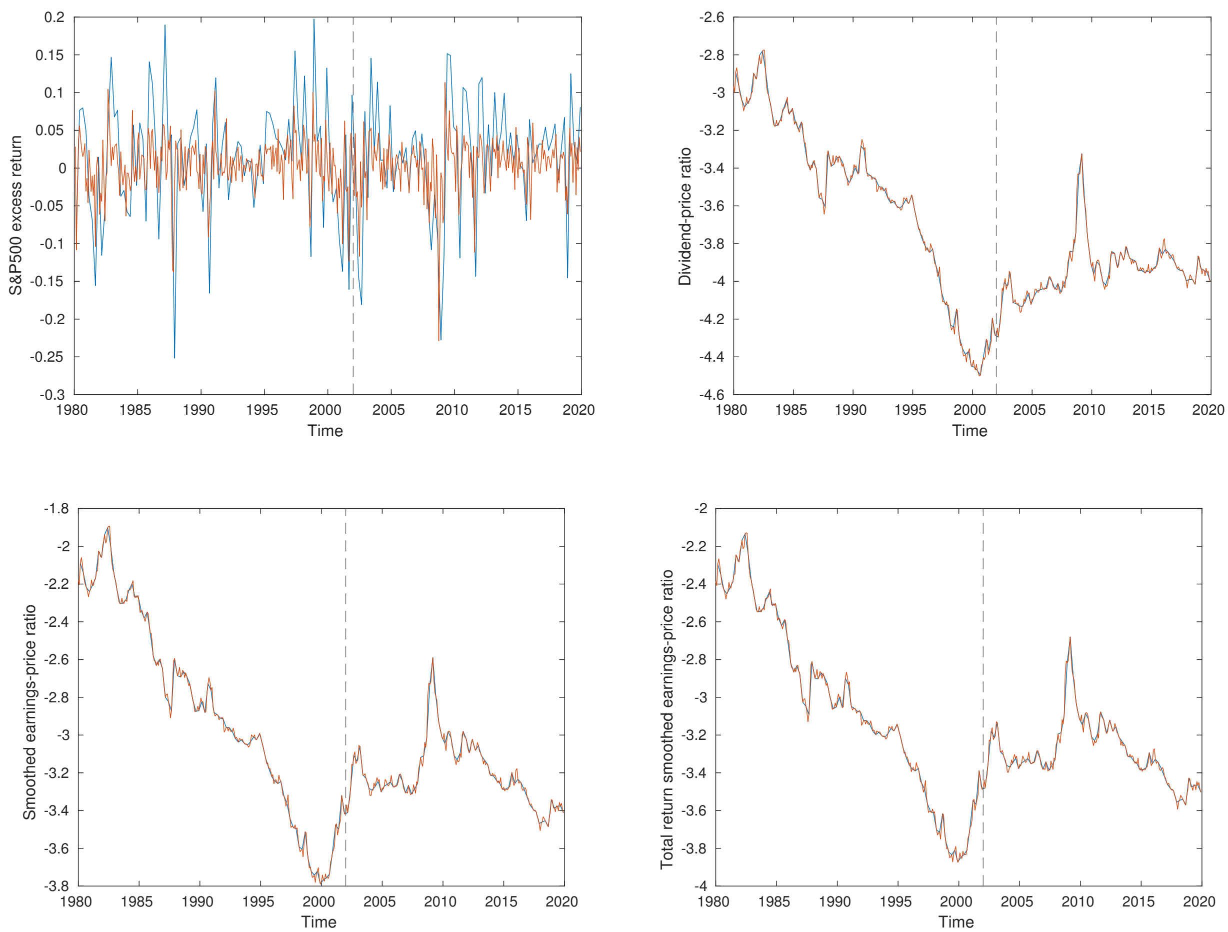
In this Section, we use our exact 10% SS-POS-based test to make inference and compare the predictive power of the valuation ratios (dividend-price ratio, smoothed earnings-price ratio, and total return smoothed earnings-price ratio) on stock market returns. The smoothed earnings-price ratio is proposed by Campbell and Shiller (1988, 2001) upon observing numerous spikes in the plot of the earnings-price ratio that had not been observed in the dividend-price ratio. The spikes were explained to be caused by recessions, which temporarily suppress corporate earnings. The latter measure is the ratio of the ten-year moving average of real earnings to current real prices and is said to possess better forecasting powers. Furthermore, the total return smoothed earnings-price ratio is recently incorporated in forecasting, as a consequence of the changes in corporate payout policy documented by Bunn et al. (2014) and Jivraj and Shiller (2017). Share repurchases (as opposed to dividends) have become the dominant approach for distributing cash to shareholders in the U.S. which may impact the smoothed earnings-price ratio through changes in growth of earnings per share. The total return smoothed earnings-price ratio corrects for this bias by reinvesting the dividends into the price index, such that the earnings per share is appropriately scaled.

6.1.1 Data description

Our data consists of monthly and quarterly observations of the aggregate S&P500 composite index for the period spanning from March 1980 to December 2019 for a total of 480 trading months or 160 trading quarters. We consider the logarithmic returns on the S&P500 in excess of the 30-day and 90-day T-bill rate. The valuation ratios under consideration are: dividend-price ratio, smoothed earnings-price ratio, and total return smoothed earnings-price ratio. The nominal monthly and quarterly prices of the value-weighted S&P500 composite index, as well as the corresponding dividends and earnings are obtained from a database provided on Robert Shiller's website. The 30-day and 90-day Treasury bill returns, on the other hand, have been retrieved from the Center for Research in Security Prices (CRSP).

in different At first glance figure 7 suggests that the predictors under consideration are highly persistent and potentially non-stationary. This visual assessment is confirmed in table 2, which presents the test statistics for the augmented Dickey-Fuller test (ADF hereafter) for all the time series. Evidently, for the full sample and the two sub-periods we fail to reject the null hypothesis of nonstationarity. The testing procedure entails estimating and testing the model in its most general form using more deterministic components than the hypothesized DGP (i.e. including both an intercept and a trend), and following Phillips and Perron (1988) sequential testing strategy thereafter, eliminating the unnecessary nuisance parameters in the process. At each stage, if the null hypothesis of orthogonality is rejected, we conclude that the model is correctly specified and that the process is stationary. Otherwise, the test is performed on a more restricted model. This procedure is continued until we arrive at the most basic form of the

Figure 7: Monthly and quarterly S&P500 excess stock returns, dividend-price, smoothed earnings-price and total return smoothed earnings-price ratios.



Note: The data spans from March 1980 to December 2019 for a total of 480 trading months and 160 trading quarters respectively. The red and the blue lines in turn correspond to the quarterly and monthly samples. To assess the predictability power of the valuation ratios, we further consider two sub-periods separated by the dashed line: one spanning from March 1980 to January 2002 and another in the period of January 2002 to January 2019.

model (with no intercept or a trend), or until the null hypothesis of unit root is rejected. As it is evident, all valuation ratios reject the null hypothesis of non-stationarity at the 5% level.

6.1.2 Predictability results

The projection technique based on the proposed 10% SS-POS test is used to build simultaneous confidence sets for the parameters of the regressions of the excess returns against the dividend-price ratio, smoothed earnings-price ratio of Campbell and Shiller (1988) and the total return smoothed earnings-price ratio of Bunn et al. (2014) and Jivraj and Shiller (2017) respectively. The results for different sub-periods and the full sample are reported in table 3. As explained in Section 4, each simultaneous confidence set is obtained by collecting all pairs of (β_0, β_1) that are not rejected using our 10% SS-POS test. Thus, a grid search is applied over an appropriate range[1] and 95% level confidence sets are constructed by retaining all the pairs (β_0, β_1) that are not rejected by the 10% SS-POS test. Alternatively, the simulated annealing algorithm can be used to solve the optimization problem (15) for each parameter β_i.

The 95% confidence intervals for the parameters β_0 and β_1 contain zero for the regressions of the excess returns against all the predictors using the T-test based on White (1980) for all periods in our study. However, using the 10% SS-POS based test, there is evidence of predictability in quarterly data in favor of all predictors for the period spanning from January 2002 to January 2019. Our findings are in line with those of Campbell and Yogo (2006) who do not find any evidence of predictability in favor of any of the predictors in the period spanning from 1952-2002.

7 Conclusion

In this chapter, we proposed simple point-optimal sign-based tests for inference in linear and non-linear predictive regression models in the presence of stochastic (or fixed) regressors. One motivation of the paper is to build valid (control the size whatever the sample size) tests for linear and non-linear predictability of stock returns. The most popular predictors of stock returns (e.g. dividend-price ratio, earning-price ratio, etc.) are known to be persistent with residuals that are correlated with the shock in the stock returns. This makes the classical predictability tests not valid, especially when the sample size is small or moderate. In addition, the proposed sign-based tests are exact, distribution-free, and robust against heteroskedasticity of unknown form and allow for serial (non-linear). Additionally, they may be inverted to build confidence regions for the parameters of the regression function. Since the point-optimal sign tests depend on the alternative hypothesis, an adaptive approach based on the split-sample technique was suggested in order to choose the appropriate alternative that controls the size and maximizes the power.

We presented a Monte Carlo study to assess the performance of the proposed "quasi"-point-optimal sign test by comparing its size and power to those of certain existing tests which are supposed to be

[1]See Section 4.1.

Table 2: Results of the ADF test on the real and nominal time-series using the general-to-specific sequential testing procedure

Series	Obs.	Predictor	p	$\delta + \mu$	μ	None
Panel A: 1980-2002						
Monthly	264	$r_t^m - r_t^f$	1	-10.959^{***}	$--$	$--$
		d/p_t	2	-2.217	-0.657	2.026
		$e/p_t^{'}$	2	-2.248	-1.171	1.721
		$e/p_t^{''}$	2	-2.160	-1.376	1.544
Quarterly	88	$r_t^m - r_t^f$	0	-9.026^{***}	$--$	$--$
		d/p_t	0	-2.209	-0.777	1.830
		$e/p_t^{'}$	0	-1.816	-1.210	1.576
		$e/p_t^{''}$	0	-1.669	-1.400	1.391
Panel B: 2002-2019						
Monthly	215	$r_t^m - r_t^f$	0	-11.369^{***}	$--$	$--$
		d/p_t	1	-2.853	-2.983^{**}	$--$
		$e/p_t^{'}$	1	-2.317	-1.938	-0.027
		$e/p_t^{''}$	1	-2.389	-1.935	0.009
Quarterly	72	$r_t^m - r_t^f$	0	-7.513^{***}	$--$	$--$
		d/p_t	1	-3.261^{*}	-3.278^{**}	$--$
		$e/p_t^{'}$	0	-2.374	-1.915	-0.095
		$e/p_t^{''}$	0	-2.448	-1.901	-0.057
Panel C: 1980-2019						
Monthly	479	$r_t^m - r_t^f$	1	-14.347^{***}	$--$	$--$
		d/p_t	2	-1.861	-2.104	0.935
		$e/p_t^{'}$	2	-1.802	-2.042	1.136
		$e/p_t^{''}$	2	-1.965	-2.161	1.056
Quarterly	160	$r_t^m - r_t^f$	0	-11.848^{***}	$--$	$--$
		d/p_t	0	-1.762	-2.051	0.876
		$e/p_t^{'}$	0	-1.732	-1.995	1.084
		$e/p_t^{''}$	0	-1.897	-2.114	0.998

Note: This table reports the results of the ADF test on the time-series in the predictive regression model. The appraoch involves using the general-to-specific sequential testing procedure to test the null hypothesis of non-stationarity, where the general form of the model is:

$$\Delta x_t = \rho x_{t-1} + \sum_{i=1}^{p-1} \psi_i \Delta x_{t-i} + \mu + \delta t + u_t \quad u_t \sim IID(0, \sigma^2) \qquad .$$

The corresponding test statistics are reported in turn for the general form of the model (including the trend δ and intercept c), the more restrictive form constituting only of an intercept c, and the case where neither the trend nor the intercept are present. The variables are defined as follows: $r_t^m - t_t^f$ are the excess logarithmic stock returns, d/p_t is the dividend-price ratio, e/p_t' is the smoothed earnings-price ratio and e/p_t'' is the total return smoothed earnings-price ratio respectively. The statistics with three asterisks (***), two asterisks (**) and one asterisk (*) are significant at the 1%, 5%. and the 10% levels respectively.

Table 3: Predictability results for the dividend-price, earnings-price and the smoothed earnings-price ratios

Series	Predictor	$\hat{\beta}$	95% confidence interval	
			10% SS-POST	WT-test
Panel A: 1980-2002				
Monthly	d/p_t	0.002	$[-0.024, 0.036]$	$[-0.008, 0.011]$
	$e/p_t^{'}$	-0.001	$[-0.044, 0.046]$	$[-0.009, 0.008]$
	$e/p_t^{''}$	-0.001	$[-0.052, 0.049]$	$[-0.010, 0.010]$
Quarterly	d/p_t	0.009	$[-0.104, 0.106]$	$[-0.028, 0.047]$
	$e/p_t^{'}$	0.003	$[-0.116, 0.104]$	$[-0.029, 0.036]$
	$e/p_t^{''}$	0.004	$[-0.126, 0.104]$	$[-0.033, 0.040]$
Panel B: 2002-2019				
Monthly	d/p_t	0.019	$[-0.220, 0.330]$	$[-0.015, 0.053]$
	$e/p_t^{'}$	0.012	$[-0.079, 0.191]$	$[-0.018, 0.042]$
	$e/p_t^{''}$	0.010	$[-0.080, 0.180]$	$[-0.021, 0.040]$
Quarterly	d/p_t	0.119	$[\mathbf{0.159}, \mathbf{0.899}]$	$[-0.001, 0.238]$
	$e/p_t^{'}$	0.089	$[\mathbf{0.042}, \mathbf{0.632}]$	$[-0.018, 0.197]$
	$e/p_t^{''}$	0.084	$[\mathbf{0.058}, \mathbf{0.697}]$	$[-0.026, 0.194]$
Panel C: 1980-2019				
Monthly	d/p_t	0.002	$[-0.041, 0.069]$	$[-0.006, 0.010]$
	$e/p_t^{'}$	0.0003	$[-0.021, 0.049]$	$[-0.007, 0.007]$
	$e/p_t^{''}$	0.0001	$[-0.039, 0.061]$	$[-0.008, 0.008]$
Quarterly	d/p_t	0.136	$[-0.094, 0.146]$	$[-0.017, 0.044]$
	$e/p_t^{'}$	0.008	$[-0.099, 0.121]$	$[-0.020, 0.036]$
	$e/p_t^{''}$	0.009	$[-0.113, 0.147]$	$[-0.023, 0.041]$

Note: This table presents the coefficient estimates, as well as the 95% confidence intervals for the variables considered in our study, by inverting the proposed 10% SS-POS-based tests and the T-test based on White (1980) variance correction. The alternatives for the 10% SS-POS tests are obtained by running OLS regressions of the excess returns against the dividend-price, smoothed earnings-price and the total return smoothed earnings-price ratios. The regressions assume the form

$$r_t^m - r_t^f = \beta_0 + \beta_1 x_{t-1} + \varepsilon_t \tag{19}$$

where r_t is the ex-post excess returns and x_{t-1} is the ex-ante predictor. The projection-based 95% confidence intervals for the 10% SS-POS tests are obtained by testing $H_0(\beta^*) : \beta = \beta^*$ on a grid for $\beta^* = (\beta_0^*, \beta_1^*)$, where the grid dimension is found by solving the optimization problem (15) for each parameter β_0 and β_1 using the simulated annealing algorithm, and consequently equally dividing each interval and finding their Cartesian product. The intervals in bold do not contain the value of zero and imply significance at the 5% level.

robust against heteroskedasticity. We considered different DGPs to illustrate different contexts that one can encounter in practice. The results show that the 10% split-sample point-optimal sign test is more powerful than the T-test, Campbell and Dufour (1995) sign-based test, and the T-test based on White (1980) variance correction.

Finally, the proposed tests were used to assess the predictive power of some financial predictors, such as the dividend-price ratio, earnings-price ratio and the smoothed earnings-price ratio of Campbell and Shiller (1988, 2001) on the annualized monthly excess stock returns. Our study suggests predictability in favor of all the predictors for the quarterly data in the period spanning from 2002 to 2009. which is consistent with the findings of Campbell and Yogo (2006), Our findings are in line with those of Campbell and Yogo (2006) who do not find any evidence of predictability in favor of any of the predictors in the period spanning from 1952-2002.

References

Banerjee, A., Dolado, J. J., Galbraith, J. W., Hendry, D., et al. (1993). Co-integration, error correction, and the econometric analysis of non-stationary data. *OUP Catalogue*.

Bunn, O., Staal, A., Zhuang, J., Lazanas, A., Ural, C., and Shiller, R. (2014). Es-cape-ing from overvalued sectors: Sector selection based on the cyclically adjusted price-earnings (cape) ratio. *The Journal of Portfolio Management*, 41(1):16–33.

Campbell, B. and Dufour, J.-M. (1995). Exact nonparametric orthogonality and random walk tests. *The Review of Economics and Statistics*, 77(1):1–16.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*, 1(3):195–228.

Campbell, J. Y. and Shiller, R. J. (2001). Valuation ratios and the long-run stock market outlook: An update. Technical report, National bureau of economic research.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.

Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of financial economics*, 81(1):27–60.

Cavanagh, C. L., Elliott, G., and Stock, J. H. (1995). Inference in models with nearly integrated regressors. *Econometric theory*, 11(5):1131–1147.

Coudin, E. and Dufour, J.-M. (2009). Finite-sample distribution-free inference in linear median regressions under heteroscedasticity and non-linear dependence of unknown form. *The Econometrics Journal*, 12:S19–S49.

Dufour, J., Jouneau, F., and Torrès, O. (2008). On (non)-testability: Applications to linear and nonlinear semiparametric and nonparametric regression models. Technical report, Technical report, Department of Economics and CIREQ, McGill University, and ....

Dufour, J.-M. and Iglesias, E. M. (2008). Finite sample and optimal adaptive inference in possibly nonstationary general volatility models with gaussian or heavy-tailed errors. Technical report, Working paper, University of Montreal.

Dufour, J.-M. and Jasiak, J. (2001). Finite sample limited information inference methods for structural equations and models with generated regressors. *International Economic Review*, 42(3):815–844.

Dufour, J.-M. and Taamouti, A. (2010). Exact optimal inference in regression models under heteroskedasticity and non-normality of unknown form. *Computational Statistics & Data Analysis*, 54(11):2532–2553.

Dufour, J.-M. and Taamouti, M. (2003). Point-optimal instruments and generalized anderson-rubin procedures for nonlinear models. Technical report, Technical report, CRDE, Université de Montréal.

Dufour, J.-M. and Torrès, O. (1998). Union-intersection and sample-split methods in econometrics with applications to ma and sure models. *Statistics Textbooks and Monographs*, 155:465–506.

Elliott, G., Müller, U. K., and Watson, M. W. (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica*, 83(2):771–811.

Elliott, G. and Stock, J. H. (1994). Inference in time series regression when the order of integration of a regressor is unknown. *Econometric theory*, 10(3-4):672–700.

Evans, G. and Savin, N. (1981). The calculation of the limiting distribution of the least squares estimator of the parameter in a random walk model. *The Annals of Statistics*, pages 1114–1118.

Evans, G. and Savin, N. (1984). Testing for unit roots: 2. *Econometrica (pre-1986)*, 52(5):1241.

Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of financial economics*, 22(1):3–25.

Goffe, W. L., Ferrier, G. D., and Rogers, J. (1994). Global optimization of statistical functions with simulated annealing. *Journal of econometrics*, 60(1-2):65–99.

Golez, B. and Koudijs, P. (2018). Four centuries of return predictability. *Journal of Financial Economics*, 127(2):248–263.

Jansson, M. and Moreira, M. J. (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica*, 74(3):681–714.

Jivraj, F. and Shiller, R. J. (2017). The many colours of cape.

Joe, H. (2014). *Dependence modeling with copulas*. Chapman and Hall/CRC.

Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies*, 28(5):1506–1553.

Lehmann, E. L. and Romano, J. P. (2006). *Testing statistical hypotheses*. Springer Science & Business Media.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–235.

Magdalinos, T. and Phillips, P. C. (2009). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory*, 25(2):482–526.

Mankiw, N. G. and Shapiro, M. D. (1986). Do we reject too often?: Small sample properties of tests of rational expectations models. *Economics Letters*, 20(2):139–145.

Maronna, R. A., Martin, R. D., Yohai, V. J., and Salibián-Barrera, M. (2019). *Robust statistics: theory and methods (with R)*. John Wiley & Sons.

Nelson, C. R. and Kim, M. J. (1993). Predictable stock returns: The role of small sample bias. *The Journal of Finance*, 48(2):641–661.

Oh, D. H. and Patton, A. J. (2016). High-dimensional copula-based distributions with mixed frequency data. *Journal of Econometrics*, 193(2):349–366.

Phillips, P. C. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica*, 82(3):1177–1195.

Phillips, P. C. and Lee, J. H. (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics*, 177(2):250–264.

Phillips, P. C. and Lee, J. H. (2016). Robust econometric inference with mixed integrated and mildly explosive regressors. *Journal of Econometrics*, 192(2):433–450.

Phillips, P. C., Magdalinos, T., et al. (2009). Econometric inference in the vicinity of unity. *Singapore Management University, CoFie Working Paper*, 7.

Phillips, P. C. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346.

Rozeff, M. S. (1984). Dividend yields are equity risk premiums. *Journal of Portfolio management*, pages 68–75.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

Stambaugh, R. F. (1985). *Bias in regressions with lagged stochastic regressors*. Center for Research in Security Prices, Graduate School of Business ....

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3):375–421.

Taamouti, A., Bouezmarni, T., and El Ghouch, A. (2014). Nonparametric estimation and inference for conditional density based granger causality measures. *Journal of Econometrics*, 180(2):251–264.

Torous, W., Valkanov, R., and Yan, S. (2004). On predicting stock returns with nearly integrated explanatory variables. *The Journal of Business*, 77(4):937–966.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *econometrica*, 48(4):817–838.

8 Appendix: Proofs

Proof of Theorem 1. From Assumption (2), the following two equalities are derived

$$P[\varepsilon_t \geq 0 \mid X] = \mathbb{E}\left(P[\varepsilon_t \geq 0 \mid \boldsymbol{\varepsilon_{t-1}}, X]\right) = \frac{1}{2}$$

with

$$\boldsymbol{\varepsilon}_0 = \{\emptyset\}, \quad \boldsymbol{\varepsilon}_{t-1} = \{\varepsilon_1, \cdots, \varepsilon_{t-1}\}, \quad \text{for} \quad t \geq 2$$

and

$$P[\varepsilon_t \geq 0 \mid \underset{\sim}{\mathrm{S}}^{\varepsilon}_{t-1}, X] = P[\varepsilon_t \geq 0 \mid \boldsymbol{\varepsilon_{t-1}}, X] = \frac{1}{2},$$

with

$$\underset{\sim}{\mathrm{S}}^{\varepsilon}_0 = \{\emptyset\}, \quad \underset{\sim}{\mathrm{S}}^{\varepsilon}_{t-1} = \{s(\varepsilon_1) = s_1, ..., s(\varepsilon_{t-1}) = s_{t-1}\}, \text{ for } t \geq 2,$$

We define the vector of signs $U(n) = (s(y_1), \cdots, s(y_n))'$, where $s(y_t) = \mathbb{1}_{\mathbb{R}^+ \cup 0}\{y_t\}$. Thus, the likelihood function of the sample in terms of signs under the null hypothesis is

$$\begin{aligned}
L(U(n), 0) &= P[s(y_1) = s_1, \cdots, s(y_n) = s_n \mid X] \\
&= P[s(\varepsilon_1) = s_1, \cdots, s(\varepsilon_n) = s_n \mid X] \\
&= \prod_{t=1}^{n} P[\varepsilon_t \geq 0 \mid \boldsymbol{\varepsilon}_{t-1}, X]^{s(\varepsilon_t)} \left(1 - P[\varepsilon_t \geq 0 \mid \boldsymbol{\varepsilon}_{t-1}, X]\right)^{1-s(\varepsilon_t)} \\
&= \prod_{t=1}^{n} \left(\frac{1}{2}\right)^{s(\varepsilon_t)} \left(1 - \frac{1}{2}\right)^{1-s(\varepsilon_t)} \\
&= \left(\frac{1}{2}\right)^{n}
\end{aligned}$$

Hence, it can be concluded that conditional on X and under the null hypothesis of orthogonality $s(y_1), \cdots, s(y_n) \overset{i.i.d}{\sim} Bi(1, 0.5)$. ■

Proof of Proposition 1. The likelihood function of sample in terms of signs $s(y_1), ..., s(y_n)$

$$L\left(U(n), \beta\right) = P\left[s(y_1) = s_1, ..., s(y_n) = s_n \mid X\right] = \prod_{t=1}^{n} P\left(s(y_t) = s_t \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right),$$

for

$$\underset{\sim}{\mathrm{S}}_0 = \{\emptyset\}, \quad \underset{\sim}{\mathrm{S}}_{t-1} = \{s(y_1) = s_1, ..., s(y_{t-1}) = s_{t-1}\}, \text{ for } t \geq 2,$$

and

$$P\left[s(y_1) = s_1 \mid \underset{\sim}{\mathrm{S}}_0, X\right] = P\left[s(y_1) = s_1 \mid X\right],$$

where each s_i, for $1 \leq t \leq n$, takes two possible values 0 and 1. According to model (1) and assumption

(2), under the null hypothesis the signs $s(y_1), \cdots, s(y_n)$ are i.i.d according to $Bi(1, 0.5)$,

$$P\left[s(y_t) = 1 \mid X\right] = P\left[s(y_t) = 0 \mid X\right] = \frac{1}{2}, \text{ for } t = 1, ..., n.$$

Consequently, under H_0

$$L_0\left(U(n), 0\right) = \prod_{t=1}^{n} P\left[s(y_t) = s_t \mid X\right] = \left(\frac{1}{2}\right)^n$$

and under H_1 we have

$$L_1\left(U(n), \beta_1\right) = \prod_{t=1}^{n} P\left[s(y_t) = s_t \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]$$

where now, for $t = 1, ..., n$,

$$y_t = \beta_1' x_{t-1} + \varepsilon_t$$

The log-likelihood ratio is given by

$$\ln\left\{\frac{L_1\left(U(n), \beta_1\right)}{L_0\left(U(n), 0\right)}\right\} = \sum_{t=1}^{n} \ln\left\{P\left[s(y_t) = s_t \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]\right\} - \mathrm{n}\ln\left\{\frac{1}{2}\right\}.$$

According to Neyman-Pearson lemma [see e.g. Lehmann (1959), page 65], the best test to test H_0 against H_1, based on $s(y_1), ..., s(y_n)$, rejects H_0 when

$$\ln\left\{\frac{L_1\left(U(n), \beta_1\right)}{L_0\left(U(n), 0\right)}\right\} \geq c$$

or when

$$\sum_{t=1}^{n} \ln\left\{P\left[s(y_t) = s_t \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]\right\} \geq c,$$

The critical value, say c, is given by the smallest constant c such that

$$P\left(\ln\left\{\frac{L_1\left(U(n), \beta_1\right)}{L_0\left(U(n), 0\right)}\right\} > c \mid H_0\right) \leq \alpha.$$

Notice that, for $t = 1, ...n$,

$$P\left[s(y_t) = s_t \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right] = P\left[y_t \geq 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]^{s(y_t)} P\left[y_t < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]^{(1-s(y_t))}, \text{ for } t = 1, ...n. \quad (20)$$

From (20), we have

$$\begin{aligned}
\ln\left\{\prod_{t=1}^{n} P\left[s(y_t)=s_t \mid \underline{\mathrm{S}}_{t-1}, X\right]\right\} &= \ln\left\{\prod_{t=1}^{n} P[y_t \geq 0 \mid \underline{\mathrm{S}}_{t-1}, X]^{s(y_t)} P[y_i < 0 \mid \underline{\mathrm{S}}_{t-1}, X]^{(1-s(y_t))}\right\} \\
&= \sum_{t=1}^{n} s(y_t) \ln\left\{P\left[y_t \geq 0 \mid \underline{\mathrm{S}}_{t-1}, X\right]\right\} \\
&\quad + \sum_{t=1}^{n} (1-s(y_t)) \ln\left\{P[y_t < 0 \mid \underline{\mathrm{S}}_{t-1}, X]\right\}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{t=1}^{n} s(y_t) \ln\left\{P\left[y_t \geq 0 \mid \underline{\mathrm{S}}_{t-1}, X\right]\right\} + \sum_{t=1}^{n} \ln\left\{P[y_t < 0 \mid \underline{\mathrm{S}}_{t-1}, X]\right\} \\
&\quad - \sum_{t=1}^{n} s(y_t) \ln\left\{P[y_t < 0 \mid \underline{\mathrm{S}}_{t-1}, X]\right\} \\
&= \sum_{t=1}^{n} s(y_t) \ln\left\{\frac{P\left[y_t \geq 0 \mid \underline{\mathrm{S}}_{t-1}, X\right]}{P\left[y_t < 0 \mid \underline{\mathrm{S}}_{t-1}, X\right]}\right\} + \sum_{t=1}^{n} \ln\left\{P[y_t < 0 \mid \underline{\mathrm{S}}_{t-1}, X]\right\}
\end{aligned}$$

Thus, the best test to test H_0 against H_1, based on $s(y_1), ..., s(y_n)$, rejects H_0 when

$$\ln\left\{\frac{L_1\left(U(n), \beta_1\right)}{L_0\left(U(n), 0\right)}\right\} = \sum_{t=1}^{n} s(y_t) \ln\left\{\frac{P\left[y_t \geq 0 \mid \underline{\mathrm{S}}_{t-1}, X\right]}{P\left[y_t < 0 \mid \underline{\mathrm{S}}_{t-1}, X\right]}\right\} + \sum_{t=1}^{n} \ln\left\{P[y_t < 0 \mid \underline{\mathrm{S}}_{t-1}, X]\right\} - \mathrm{n} \ln\left\{\frac{1}{2}\right\} \geq c$$

or when

$$\ln\left\{\frac{L_1\left(U(n), \beta_1\right)}{L_0\left(U(n), 0\right)}\right\} = \sum_{t=1}^{n} s(y_t) \ln\left\{\frac{P\left[y_t \geq 0 \mid \underline{\mathrm{S}}_{t-1}, X\right]}{P\left[y_t < 0 \mid \underline{\mathrm{S}}_{t-1}, X\right]}\right\} \geq c_1(\beta_1)$$

where the critical value $c_1(\beta_1)$ is chosen so that

$$P\left[S_n(\beta_1) > c_1(\beta_1) \mid H_0\right] \leq \alpha$$

α is an arbitrary significance level. ■

Proof of Assumption 1.

Let $y_1, \cdots, y_t$ be linearly explained by a vector variable x_t

$$y_t = \beta_1' x_{t-1} + \varepsilon_t$$

and suppose y_t follows a Markov process of order one, such that

$$y_t, y_{t-1} \mid X \sim N\left(\underbrace{\begin{bmatrix} \beta' x_{t-1} \\ \beta' x_{t-2} \end{bmatrix}}_{\boldsymbol{\mu}}, \underbrace{\begin{bmatrix} \sigma^2_{\varepsilon_t} & \sigma_{\varepsilon_t \varepsilon_{t-1}} \\ \sigma_{\varepsilon_{t-1}\varepsilon_t} & \sigma^2_{\varepsilon_{t-1}} \end{bmatrix}}_{\Sigma}\right), \quad t = 2, \cdots, n.$$

where it is assumed, without loss of generality, that ε_t and ε_{t-1} are serially correlated with $\sigma_{\varepsilon_{t-1}\varepsilon_t} \neq 0$. Then the signs $s(y_1), \cdots, s(y_t)$ are Bernoulli variables with conditional joint distributions fully determined by $P_{s(y_t)|X}$, $P_{s(y_{t-1})|X}$, and either $P_{s(y_t)|s(y_{t-1}),X}$ or $P_{s(y_t),s(y_{t-1})|X}$, where

$$\begin{aligned} P_{s(y_t)|s(y_{t-1}),X} &:= P[s(y_t) = 1 \mid s(y_{t-1}) = 1, X] \\ P_{s(y_t),s(y_{t-1})|X} &:= P[s(y_t) = 1, s(y_{t-1}) = 1 \mid X], \end{aligned}$$

which may alternatively be expressed as

$$\begin{aligned} P[s(y_t) = 1 \mid s(y_{t-1}) = 1, X] &= P[y_t \geq 0 \mid y_{t-1} \geq 0 \mid X] = P[\varepsilon_t \geq -\beta' x_{t-1} \mid \varepsilon_{t-1} \geq -\beta' x_{t-2}, X] \\ P[s(y_t) = 1, s(y_{t-1}) = 1 \mid X] &= P[y_t \geq 0, y_{t-1} \geq 0 \mid X] = P[\varepsilon_t \geq -\beta' x_{t-1}, \varepsilon_{t-1} \geq -\beta' x_{t-2} \mid X]. \end{aligned}$$

As the dependence in the pair-wise probabilities is determined by the covariance matrix Σ, with $\sigma_{\varepsilon_t,\varepsilon_{t-1} \neq 0}$, this in turn implies that the signs $s(y_t), s(y_{t-1})$ for $t = 1, \cdots, n$ are dependent and follow a Markov process of order one. These findings are extended to the case where the signs exhibit non-linear serial dependence.

■

Proof of Corollary 1. From test statistic $S_n(\beta_1)$ in Proposition 1 and under assumption **A1**, we have:

$$\begin{aligned} \tilde{S}_n(\beta_1) &= \sum_{t=1}^{n} s(y_t) \ln \left\{ \frac{P\left[y_t \geq 0 \mid \underset{\sim}{S}_{t-1}, X\right]}{P\left[y_t < 0 \mid \underset{\sim}{S}_{t-1}, X\right]} \right\} \\ &= \sum_{t=1}^{n} s(y_t) \left\{ \ln \left\{ P\left[y_t \geq 0 \mid \underset{\sim}{S}_{t-1}, X\right] \right\} - \ln \left\{ P\left[y_t < 0 \mid \underset{\sim}{S}_{t-1}, X\right] \right\} \right\} \\ &= \sum_{t=1}^{n} s(y_t) \left\{ \begin{array}{c} \ln \left\{ P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]^{s(y_{t-1})} P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]^{1-s(y_{t-1})} \right\} \\ - \ln \left\{ P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right]^{s(y_{t-1})} P\left[y_t < 0 \mid y_{t-1} < 0, X\right]^{1-s(y_{t-1})} \right\} \end{array} \right\} \\ &= \sum_{t=1}^{n} s(y_t) \left\{ \begin{array}{c} s(y_{t-1}) \ln \left\{ P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right] \right\} + (1 - s(y_{t-1})) \ln \left\{ P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right] \right\} \\ -s(y_{t-1}) \ln \left\{ P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right] \right\} - (1 - s(y_{t-1})) \ln \left\{ P\left[y_t < 0 \mid y_{t-1} < 0, X\right] \right\} \end{array} \right\} \end{aligned}$$

Observe that:

$$\begin{aligned}
\ln\left\{\frac{P\left[y_t \geq 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}{P\left[y_t < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}\right\} &= \ln\left\{P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]^{s(y_{t-1})} P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]^{1-s(y_{t-1})}\right\} \\
&\quad - \ln\left\{P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right]^{s(y_{t-1})} P\left[y_t < 0 \mid y_{t-1} < 0, X\right]^{1-s(y_{t-1})}\right\} \\
&= s(y_{t-1}) \ln\left\{P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]\right\} \\
&\quad + (1 - s(y_{t-1})) \ln\left\{P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]\right\} \\
&\quad - s(y_{t-1}) \ln\left\{P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right]\right\} \\
&\quad - (1 - s(y_{t-1})) \ln\left\{P\left[y_t < 0 \mid y_{t-1} < 0, X\right]\right\}
\end{aligned}$$

$$\begin{aligned}
&= s(y_{t-1}) \ln\left\{P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]\right\} + \ln\left\{P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]\right\} \\
&\quad - s(y_{t-1}) \ln\left\{P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]\right\} - s(y_{t-1}) \ln\left\{P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right]\right\} \\
&\quad - \ln\left\{P\left[y_t < 0 \mid y_{t-1} < 0, X\right]\right\} + s(y_{t-1}) \ln\left\{P\left[y_t < 0 \mid y_{t-1} < 0, X\right]\right\}
\end{aligned}$$

$$\begin{aligned}
&= s(y_{t-1}) \left\{\ln\left\{\frac{P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]}{P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right]}\right\} - \ln\left\{\frac{P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]}{P\left[y_t < 0 \mid y_{t-1} < 0, X\right]}\right\}\right\} \\
&\quad + \ln\left\{\frac{P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]}{P\left[y_t < 0 \mid y_{t-1} < 0, X\right]}\right\}
\end{aligned}$$

Hence,

$$\begin{aligned}
\tilde{S}_n(\beta_1) &= \sum_{t=1}^{n} s(y_t) \ln\left\{\frac{P\left[y_t \geq 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}{P\left[y_i < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}\right\} \\
&= \sum_{t=1}^{n} s(y_t) \left\{\begin{array}{c} s(y_{t-1}) \left\{\ln\left\{\frac{P[y_t \geq 0 \mid y_{t-1} \geq 0, X]}{P[y_t < 0 \mid y_{t-1} \geq 0, X]}\right\} - \ln\left\{\frac{P[y_t \geq 0 \mid y_{t-1} < 0, X]}{P[y_t < 0 \mid y_{t-1} < 0, X]}\right\}\right\} \\ + \ln\left\{\frac{P[y_t \geq 0 \mid y_{t-1} < 0, X]}{P[y_t < 0 \mid y_{t-1} < 0, X]}\right\} \end{array}\right\} \\
&= \sum_{t=1}^{n} s(y_t) \ln\left\{\frac{P\left[y_t \geq 0 \mid y_t < 0, X\right]}{P\left[y_t < 0 \mid y_t < 0, X\right]}\right\} + \sum_{t=1}^{n} s(y_t) s(y_{t-1}) \left\{\begin{array}{c} \ln\left\{\frac{P[y_t \geq 0 \mid y_{t-1} \geq 0, X]}{P[y_t < 0 \mid y_{t-1} \geq 0, X]}\right\} \\ - \ln\left\{\frac{P[y_t \geq 0 \mid y_{t-1} < 0, X]}{P[y_t < 0 \mid y_{t-1} < 0, X]}\right\} \end{array}\right\} \\
&= \sum_{t=1}^{n} a_t \; s(y_t) + \sum_{t=1}^{n} b_t \; s(y_t) s(y_{t-1})
\end{aligned}$$

where

$$\tilde{a}_1 = \ln\left\{\frac{P\left[y_1 \geq 0 \mid X\right]}{P\left[y_1 < 0 \mid X\right]}\right\} = \ln\left\{\frac{1 - P\left[\varepsilon_1 < -\beta_1' x_0 \mid X\right]}{P\left[\varepsilon_1 < -\beta_1' x_0 \mid X\right]}\right\}$$

$$\tilde{b}_1 = \ln\left\{\frac{P\left[y_1 \geq 0 \mid X\right]}{P\left[y_1 < 0 \mid X\right]}\right\} - \ln\left\{\frac{P\left[y_1 \geq 0 \mid X\right]}{P\left[y_1 < 0 \mid X\right]}\right\} = 0$$

and for $t = 2, ..., n$

$$
\begin{aligned}
a_t &= \ln \left\{ \frac{P\left[y_t \geq 0 \mid y_t < 0, X\right]}{P\left[y_t < 0 \mid y_t < 0, X\right]} \right\}, \\
b_t &= \ln \left\{ \frac{P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]}{P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right]} \right\} - \ln \left\{ \frac{P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]}{P\left[y_t < 0 \mid y_{t-1} < 0, X\right]} \right\}.
\end{aligned}
$$

Observe that:

$$
\begin{aligned}
P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right] &= 1 - P\left[y_t < 0 \mid y_{t-1} < 0, X\right] \\
&= 1 - \frac{P\left[y_t < 0,\ y_{t-1} < 0 \mid X\right]}{P\left[y_{t-1} < 0 \mid X\right]} \\
&= 1 - \frac{P\left[\varepsilon_t < -\beta_1' x_{t-1},\ \varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}{P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}, \\
P\left[y_t < 0 \mid y_{t-1} < 0, X\right] &= \frac{P\left[y_t < 0,\ y_{t-1} < 0 \mid X\right]}{P\left[y_{t-1} < 0 \mid X\right]} \\
&= \frac{P\left[\varepsilon_t < -\beta_1' x_{t-1},\ \varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}{P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]} \\
P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right] &= 1 - P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right] \\
&= 1 - \frac{P\left[y_t < 0,\ y_{t-1} \geq 0 \mid X\right]}{P\left[\ y_{t-1} \geq 0 \mid X\right]} \\
&= 1 - \frac{P\left[y_t < 0 \mid X\right]}{P\left[\ y_{t-1} \geq 0 \mid X\right]} \left(P\left[\ y_{t-1} \geq 0 \mid y_t < 0, X\right]\right) \\
&= 1 - \frac{P\left[y_t < 0 \mid X\right]}{P\left[\ y_{t-1} \geq 0 \mid X\right]} \left(1 - P\left[\ y_{t-1} < 0 \mid y_t < 0, X\right]\right) \\
&= 1 - \left(\frac{P\left[y_t < 0 \mid X\right]}{P\left[\ y_{t-1} \geq 0 \mid X\right]} - \frac{P\left[\ y_{t-1} < 0,\ y_t < 0 \mid X\right]}{P\left[\ y_{t-1} \geq 0 \mid X\right]} \right) \\
&= 1 - \left(\frac{P\left[y_t < 0 \mid X\right]}{1 - P\left[y_{t-1} < 0 \mid X\right]} - \frac{P\left[\ y_{t-1} < 0,\ y_t < 0 \mid X\right]}{1 - P\left[y_{t-1} < 0 \mid X\right]} \right) \\
&= 1 - \left[\frac{P\left[\varepsilon_t < -\beta_1' x_{t-1} \mid X\right]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]} - \frac{P\left[\ \varepsilon_{t-1} < -\beta_1' x_{t-2},\ \varepsilon_t < -\beta_1' x_{t-1} \mid X\right]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]} \right)
\end{aligned}
$$

$$
\begin{aligned}
P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right] &= \frac{P\left[y_t < 0,\ y_{t-1} \geq 0 \mid X\right]}{P\left[y_{t-1} \geq 0 \mid X\right]} \\
&= \frac{P\left[y_{t-1} \geq 0 \mid\ \ y_t < 0, X\right] P\left[y_t < 0 \mid X\right]}{P\left[y_{t-1} \geq 0 \mid X\right]} \\
&= \frac{P\left[y_t < 0 \mid X\right]}{P\left[y_{t-1} \geq 0 \mid X\right]} \left(1 - P\left[y_{t-1} < 0 \mid\ \ y_t < 0, X\right]\right) \\
&= \frac{P\left[y_t < 0 \mid X\right]}{1 - P\left[y_t < 0 \mid X\right]} - \frac{P\left[y_{t-1} < 0,\ y_t < 0 \mid X\right]}{1 - P\left[y_t < 0 \mid X\right]} \\
&= 1 - P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]
\end{aligned}
$$

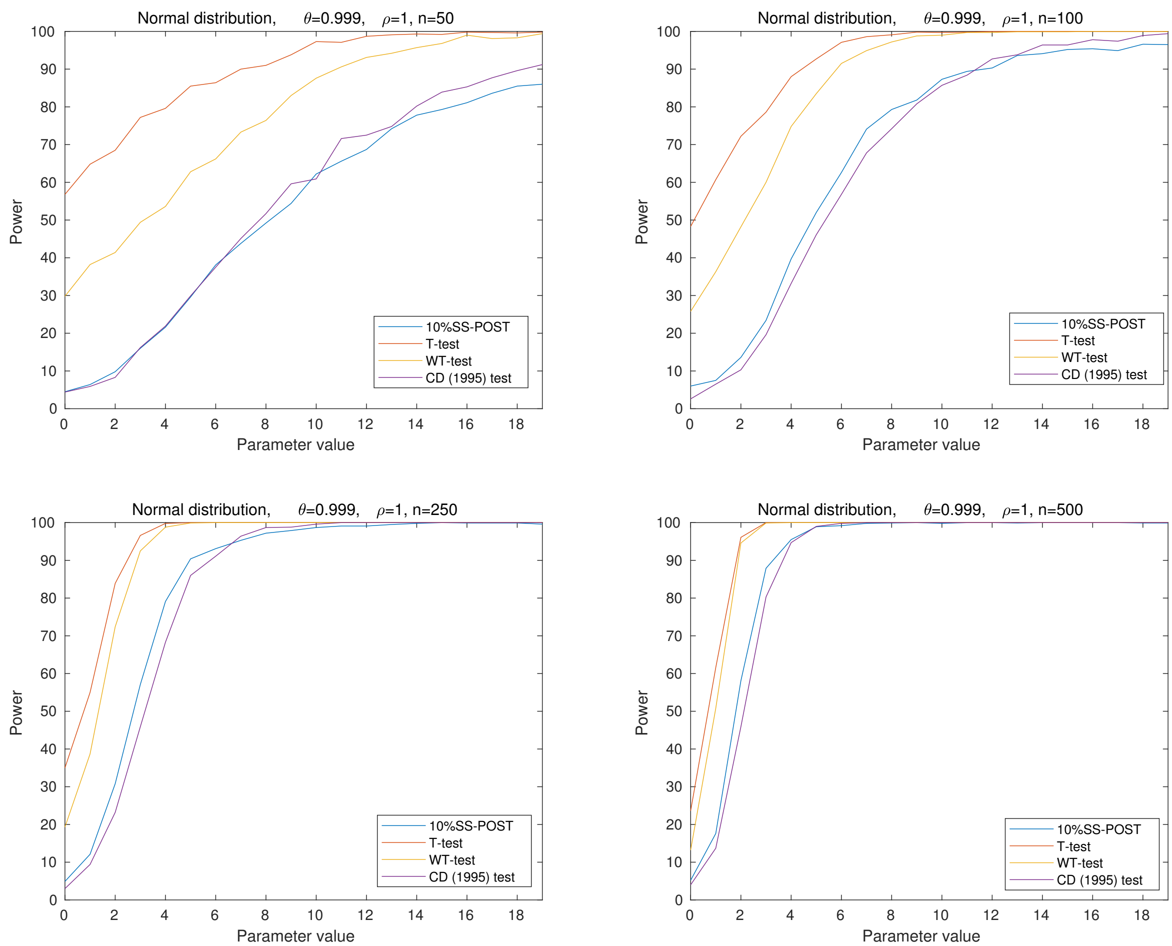
We also have:

$$
\begin{aligned}
\tilde{a}_1 &= \ln\left\{\frac{P\left[y_1 \geq 0 \mid X\right]}{P\left[y_1 < 0 \mid X\right]}\right\} \\
&= \ln\left\{\frac{1 - P\left[y_1 < 0 \mid X\right]}{P\left[y_1 < 0 \mid X\right]}\right\} \\
&= \ln\left\{\frac{1 - P\left[\varepsilon_1 < -\beta_1' x_0 \mid X\right]}{P\left[\varepsilon_1 < -\beta_1' x_0 \mid X\right]}\right\} \\
\tilde{b}_1 &= \ln\left\{\frac{P\left[y_1 \geq 0 \mid X\right]}{P\left[y_1 < 0 \mid X\right]}\right\} - \ln\left\{\frac{P\left[y_1 \geq 0 \mid X\right]}{P\left[y_1 < 0 \mid X\right]}\right\} = 0
\end{aligned}
$$

■

Additional simulations.

Figure 8: Power comparisons: different tests. Normal distributions with contemporaneous correlation of $\rho = 1$, in (18) and local-to-unity autoregression parameter $\theta = 0.999$, in (17) for different sample sizes.



Note: These figures compare the power function of the 10% SS-POS test with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

■