L'Actualité économique

Finite-Sample Sign-Based Inference in Linear and Nonlinear Regression Models with Applications in Finance

Abderrahim Taamouti

Volume 91, numéro 1-2, mars–juin 2015

Identification, Simulation and Finite-Sample Inference

URI : https://id.erudit.org/iderudit/1036915ar
DOI : https://doi.org/10.7202/1036915ar

Aller au sommaire du numéro

Éditeur(s)

HEC Montréal

ISSN

0001-771X (imprimé)
1710-3991 (numérique)

Résumé de l'article

We review several exact sign-based tests that have been recently proposed for testing orthogonality between random variables in the context of linear and nonlinear regression models. The sign tests are very useful when the data at the hands contain few observations, are robust against heteroskedasticity of unknown form, and can be used in the presence of non-Gaussian errors. These tests are also flexible since they do not require the existence of moments for the dependent variable and there is no need to specify the nature of the feedback between the dependent variable and the current and future values of the independent variable. Finally, we discuss several applications where the sign-based tests can be used to test for multi-horizon predictability of stock returns and for the market efficiency.

L'Actualité économique, Revue d'analyse économique, vol. 91, nos 1-2, mars-juin 2015

FINITE-SAMPLE SIGN-BASED INFERENCE IN LINEAR AND NONLINEAR REGRESSION MODELS WITH APPLICATIONS IN FINANCE*

Abderrahim TAAMOUTI
Durham University Business School
abderrahim.taamouti@durham.ac.uk

Abstract–We review several exact sign-based tests that have been recently proposed for testing orthogonality between random variables in the context of linear and nonlinear regression models. The sign tests are very useful when the data at the hands contain few observations, are robust against heteroskedasticity of unknown form, and can be used in the presence of non-Gaussian errors. These tests are also flexible since they do not require the existence of moments for the dependent variable and there is no need to specify the nature of the feedback between the dependent variable and the current and future values of the independent variable. Finally, we discuss several applications where the sign-based tests can be used to test for multi-horizon predictability of stock returns and for the market efficiency.

Introduction

In this chapter we survey several recent developments on sign-based inference. The literature on sign tests is not new and several books and monographs have be written on these tests in the context of i.i.d. data; see e.g. Boldin *et al.* (1997). However, the focus here is on reviewing new sign-based tests that have been proposed to test orthogonality between random variables in the context of linear and nonlinear regression models and in the presence of both independent and dependent data. We also illustrate how these tests can be used to overcome well known problems encountered when testing important financial theories.

As we know, regression errors in economic and financial data frequently exhibit non-normal distributions and heteroskedasticity. In the presence of several types of heteroskedasticity, usual "robust" tests–such as tests based on White (1980)-type variance corrections–remain plagued by poor size control and/or low power. In

* We thank the Editors Marie-Claude Beaulieu, Jean-Marie Dufour, and Lynda Khalaf and an anonymous referee for his/her very useful comments. Financial support from Durham University Business School and the Spanish Ministry of Education through grants SEJ 2007-63098 and #ECO2010-19357 are acknowledged.

addition, the available *exact parametric* tests typically assume Gaussian disturbances. The latter assumption is often unrealistic and, in the presence of heavy tails and asymmetric distributions, the associated tests may easily not perform well in terms of size control or power. Moreover, statistical procedures for inference on parameters of *nonlinear* models are typically based on asymptotic approximations, which may easily be not reliable in finite samples; see Dufour (2003).

In the last two decades a number of new sign-based test procedures have been developed in the literature to deal with the above problems. In the presence of only one explanatory variable, Campbell and Dufour (1991, 1995, 1997) and Luger (2003) propose nonparametric analogues of the t-test, based on sign and signed rank statistics, which are applicable when regressors involve feedback of the type considered by Mankiw and Shapiro (1986). These tests are exact even when the disturbances are asymmetric, non-normal, and heteroskedastic. In the presence of non-stochastic regressors, Dufour and Taamouti (2010) propose simple point-optimal sign-based tests in linear and nonlinear multivariate regressions, which are valid under non-normality and heteroskedascticity of unknown form, provided the errors have median zero conditional on the explanatory variables. The proposed tests are exact, distribution-free, and may be inverted to build confidence regions for the vector of unknown parameters. Furthermore, an important feature of these tests comes from the fact that they trace out the *power envelope*, *i.e.* the maximum achievable power for a given testing problem. The power envelope provides an obvious benchmark against which test procedures can be evaluated. Coudin and Dufour (2009) extend the work by Boldin *et al.* (1997) to account for serial dependence and discrete distributions. In particular, they develop finite-sample and distribution-free sign-based tests and confidence sets for the parameters of a linear multivariate regression model, where no parametric assumption is imposed on the noise distribution. In addition to non-normality and heteroscedasticity, their set-up allows for nonlinear serial dependence of unknown forms. To build their sign tests, they first consider a "*mediangale*" structure under which the signs of *mediangale* sequences follow a nuisance-parameter-free distribution despite the presence of non-linear dependence and heterogeneity of unknown form.

The present chapter faithfully follows the text of the abovementioned papers to survey the exact sign-based tests that have been recently proposed for testing orthogonality between random variables. These tests are very useful when the data at the hands contain few observations, and they are very flexible since they do not require the existence of moments for the dependent variable and there is no need to specify the nature of the feedback between the dependent variable and the current and future values of the independent variable.

The above statistical procedures are motivated in at least two ways. First, it is well known that hypotheses on means (moments) are not testable in nonparametric setups even under the apparently restrictive assumption that observation are independent and identically distributed (*i.i.d.*): if a test has level α for testing the null hypothesis that the mean of *i.i.d.* observations has a given value, then its power cannot be larger than the level α under any alternative of the mean; see Bahadur and Savage (1956). Similar results hold for the coefficients of regression models; see Dufour, Jouneau, and Torrès (2008). In other words, moments are not empirically meaningful in many common nonparametric models. This provides a strong reason for focusing on quantiles (such as median) in nonparametric models — instead of moments — because quantiles are not affected by such problems of nontestability. Second, in the presence of general heteroskedasticity, Lehmann and Stein (1949) and Pratt and Gibbons (1981) show that sign methods are the only possible way of producing valid inference in finite samples; see also Dufour and Hallin (1991) and Dufour (2003). If a test has level α for testing the null hypothesis that observations are independent each with a distribution symmetric about zero, then its level must be equal to α conditional on the absolute values of the observations: in other words, it must be a *sign test*. For a more detailed discussion of statistical inference impossibilities in nonparametric models, see Dufour and Hallin (1991) and Dufour (2003).

Finally, we discuss several applications where sign-based tests are used to test the multi-horizon predictability of stock returns (Liu and Maynard, 2007) and for the market efficiency (Gungor and Luger, 2009).

The plan of the chapter is as follows. In the first section, we present a general framework for reviewing different sign-based tests. In Section 2, we review many tests that have been proposed in the context of simple regression model. In Section 3, we survey several sign-based tests for testing parameters in multivariate linear and nonlinear regressions and in the presence of both independent and dependent data. In Section 4, we discuss two applications where sign-based tests are used to test the multi-horizon predictability of stock returns and market efficiency. We conclude in the laast section.

1. General Framework

In this section, we describe a general framework for reviewing several exact sign-based tests that have been recently proposed for testing orthogonality between random variables. These tests are motived in the context of the following regression

$$Y_t = \mu + f\left(X_t; \beta\right) + \varepsilon_t, \tag{1}$$

where assumptions on the error term ε_t, on the functional form $f(.)$, on the randomness and dimension of X, and on the presence or absence of the intercept μ lead to different tests. In particular, we discuss several *finite-sample* tests of independence between Y and X which are exact under weak assumptions concerning the distribution of Y and the relationship between Y and X. These tests can also differ depending on whether or not the concept of optimality (power) is under consideration.

For the first group of exact sign-based tests that we review below, the functional form $f(.)$ is assumed to be linear and X is stochastic. However, both Y and X cannot be multivariate and we simply assume that Y has median zero. No additional assumption other than the independence of Y_t with respect to the past (hereafter I_{t-1}) governs the relationship between Y and X. In other words, this first group of sign-based tests are developed within the framework of the following general specification involving the random variables $Y_1,\ldots, Y_n, X_0,\ldots, X_{n-1}$, and the corresponding information vectors $I_t = (X_0, X_1,\ldots, X_t, Y_1,\ldots,Y_t)'$, where $t = 0,\ldots,n-1$, with the convention $I_t = (X_0)$:

$$Y_t \text{ is independent of } I_{t-1}, \text{ for each } t = 1,\ldots,n, \tag{2}$$

$$P[Y_t > 0] = P[Y_t < 0], \text{for } t = 1,\ldots,n. \tag{3}$$

Assumption (2) indicates that Y_t is independent of the past values of Y_t and X_t, while assumption (3) states that $Y_1,\ldots, Y_n$ have median zero. As discussed in Campbell and Dufour (1995), these assumptions leave open the possibly of feedback from Y_t to current and future values of the X-variable without specifying the form of feedback. Furthermore, the variables Y and X may have discrete distributions, which includes the possibility of non-zero probability mass at zero; as well, the variables Y need not be Gaussian nor identically distributed. In what follows, the additional assumption that $Y_1,\ldots, Y_n$ have distributions symmetric about zero (m_0) is also considered:

$$Y_1,\ldots, Y_n \text{ have continuous distributions symmetric about zero } (m_0). \tag{4}$$

For the second group of sign-based tests that we discuss below, X can be multivariate. Furthermore, two sets of assumptions are separately considered depending on whether or not Y is assumed to be independent. When the process of Y is supposed to be independent, then it is assumed that

$$Y_1,\ldots, Y_n \text{ are independent conditional on } \mathbf{X} \tag{5}$$

and the error term in the regression (1) satisfies:

$$\mathrm{P}[\varepsilon_t > 0 \mid \mathbf{X}] = \mathrm{P}[\varepsilon_t < 0 \mid \mathbf{X}] = \frac{1}{2}, t = 1,\ldots,n, \tag{6}$$

where $\mathbf{X} = [X_1,\ldots,X_n]'$ is an $n \times k$ matrix. Note that Assumption (6) entails that the error term ε_t has no mass at zero, *i.e.* $\mathrm{P}[\varepsilon_t = 0 \mid \mathbf{X}] = 0$ for all t. Moreover, it is clear that assumptions (5) and (6) are more restrictive than assumptions (2) and (3) and (4). The former will help to build *optimal* sign-based tests which are valid for both X univariate and multivariate. Optimality here is in the Neyman-Pearson sense, thus these tests maximize the power function under the level constraint; see for example Lehmann (1959, page 65). Finally, when the process of Y is supposed to be dependent, it is mainly assumed that the the error term ε_t is a *mediangale* process, where the latter term is defined in Coudin and Dufour (2009); see also Section 3.2.

2. Sign-based Tests for Simple Regressions

In the presence of only one explanatory variable, Campbell and Dufour (1991, 1995, 1997) and Luger (2003) propose nonparametric analogues of the t-test, based on sign and signed rank statistics, which are applicable when regressors involve feedback of the type considered by Mankiw and Shapiro (1986). These tests are exact even when the disturbances are asymmetric, non-normal, and heteroskedastic. Campbell and Dufour (1991, 1995) have proposed exact sign and signed rank statistics in the absence of a nuisance parameter (drift parameter μ), and Campbell and Dufour (1997) and Luger (2003) build sign and signed rank statistics in the presence of a nuisance parameter (drift parameter μ).

2.1 *Sign-based Tests without Nuisance Parameters*

Campbell and Dufour (1991) have introduced non-parametric analogues of the *t*-test, based on sign statistics and Wilcoxon signed-rank statistics, that are applicable in the context of an important variant of the Mankiw and Shapiro (1986) model. Using Monte Carlo techniques, the latter found that the standard testing procedure, such as *t*-test, that is used to assess the rationality of expectations may be considerably greater than its nominal level in a fairly simple model and in the presence of large samples. Campbell and Dufour (1995) have considerably generalized the results in Campbell and Dufour (1991), where various nonparametric statistics are introduced to deal with a variant of the Mankiw and Shapiro (1986) model. In particular, the nature of the allowed feedback is considerably more general in Campbell and Dufour (1995) and exact distributional results are established. For these reasons, in what follows we focus on only reviewing the results in Campbell and Dufour (1995).

Campbell and Dufour (1995) consider the following linear simple regression model:

$$Y_t = \beta X_{t-1} + \varepsilon_t, \tag{7}$$

where the drift parameter is equal to zero, X is a scalar independent covariate, and the error term ε_t has the same properties as Y_t in (2) and (3) and (4). Suppose we wish to test the null hypothesis:

$$H_0 : \beta = \beta_0. \tag{8}$$

To test the null H_0 in (8), Campbell and Dufour (1995) propose to use nonparametric analogues of Student's *t*-statistics based on sign (rank) of the observations, which is derived from

$$T = \left(\hat{\beta} - \beta_0\right) / \hat{\sigma}^2 \left(\sum_{t=1}^{n} X_{t-1}^2\right)^{-1/2} = \sum_{t=1}^{n} V_t \tag{9}$$

where $\hat{\beta} = \sum_{t=1}^{n} Y_t X_{t-1} / \sum_{t=1}^{n} X_{t-1}^2$, $\hat{\sigma}^2 = \sum_{t=1}^{n} \left(Y_t - \hat{\beta} X_{t-1}\right)^2 / (n-1)$, and

$$V_t = \left(Y_t - \beta_0 X_{t-1}\right) X_{t-1} / \hat{\sigma}^2 \left(\sum_{t=1}^{n} X_{t-1}^2\right)^{1/2}.$$

The nonparametric test procedures of Campbell and Dufour (1995) abstract from the specific values of V_t to consider simply its sign and possibly the rank of its absolute value among $|V_1|,...,|V_n|$. Those procedures consider the products $Z_t = (Y_t - \beta_0 X_{t-1})X_{t-1}$ as the basic building block in the definition of various nonparametric statistics. Thus, a nonparametric analogue of the t-statistic in (9) is the sign statistic given by:

$$S_g = \sum_{t=1}^{n} u\big((Y_t - \beta_0 X_{t-1})\, g_{t-1}\big), \tag{10}$$

where

$$u(z) = \begin{cases} 1, & if\ z \geq 0 \\ 0, & if\ z < 0, \end{cases} \tag{11}$$

with $g_t = g_t(I_t)$, $t = 0,...,n-1$, is a sequence of measurable functions of the information vector I_t. The latter functions allow one to consider various transformations of the data, provided g_t depends only on past and current values of X_τ and Y_τ $(\tau \leq t)$. A special case of (10) is the following test statistic

$$S_0 = \sum_{t=1}^{n} u\big((Y_t - \beta_0 X_{t-1})X_{t-1}\big).$$

Without loss of generality, in what follows we focus on testing $H_0 : \beta = 0$, which corresponds to testing orthogonality between Y and X. The following proposition establishes the exact distribution of S_g when Y_t and g_t are continuous variables (have no probability mass at zero).

Proposition 1 (Campbell and Dufour, 1995) *Let* $Y = (Y_1,...,Y_n)'$ *and* $X = (X_0,...,X_{n-1})'$ *be two* nx1 *random vectors which satisfy assumptions (2) and (3). Suppose further that* $P[Y_t = 0] = 0$, *for* $t = 1,...,n$ *and let* $g_t = g_t(I_t)$, $t = 0,...,n-1$, *be a sequence of measurable functions of* I_t *such that* $P[g_t = 0] = 0$, *for* $t = 0,...,n-1$. *Then the statistic* S_g *defined by (10) follows a* $Bi(n,0.5)$ *distribution, i.e.*

$$P[S_g = x] = C_n^x (1/2)^n \text{ for } x = 0,1,...,n, \tag{12}$$

where $C_n^x = n!/[x!(n-x)!]$.

Assumption $P[Y_t = 0] = P[g_t = 0] = 0$ in Proposition 1 means that Y_t and g_t have no mass at zero, which holds when these variables have continuous distributions. In addition, the conditions of Proposition 1 are quite flexible since there are no assumptions concerning the existence of moments of Y_t; heteroskedasticity of unknown form is permitted; the nature of the feedback between Y_t and current and future values of X_{t+s} $(s \geq 0)$ is not specified.

Campbell and Dufour (1995) consider other test statistics that are based on sign and ranks under the further assumption in (4), with $m_0 = 0$. In particular, they consider the following signed rank statistics

$$W_g = \sum_{t=1}^{n} u(Y_t g_{t-1}) R_{1t}^{+}, \tag{13}$$

$$SR_g = \sum_{t=1}^{n} u(Y_t g_{t-1}) R_{2t}^{+}, \tag{14}$$

where R_{1t}^{+} is the rank of $|Y_t g_{t-1}|$, *i.e.* $R_{1t}^{+} = \sum_{j=1}^{n} u\left(|Y_t g_{t-1}| - |Y_j g_{j-1}|\right)$ the rank of $|Y_t g_{t-1}|$ when $|Y_1 g_0|, \dots, |Y_n g_{n-1}|$ are put in ascending order, while R_{2t}^{+} denotes the rank of $|Y_t|$ among $|Y_1|, \dots, |Y_n|$. Special cases of signed rank statistics in (13) and (14) are obtained by taking $g_t = X_t$:

$$W_0 = \sum_{t=1}^{n} u(Y_t X_{t-1}) R_{1t}^{+}, \; SR_0 = \sum_{t=1}^{n} u(Y_t X_{t-1}) R_{2t}^{+},$$

computed by weighting the sign of each positive product $Y_t X_{t-1}$ by the rank of its absolute value. As pointed out by Campbell and Dufour (1995), the possibility of feedback makes it impossible to establish in general that W_0 and W_g are distributed as a Wilcoxon signed rank variate, *i.e.*, as $W = \sum_{t=1}^{n} tB_t$ where $B_1, \dots, B_n$ are independent random variables such that $P[B_t = 0] = P[B_t = 1] = 0.5$, for $t = 1, \dots, n$. However, in the absence of feedback, Campbell and Dufour (1995) derive the following result.

Proposition 2 (Campbell and Dufour, 1995) *Let* $Y = (Y_1, \dots, Y_n)'$ *and* $X = (X_0, \dots, X_{n-1})'$ *be independent* $n \times 1$ *random vectors such that (2) and (4), for* $m_0 = 0$ *hold. Let* $g_t = g_t(X)$, $t = 0, \dots, n-1$, *be a sequence of measurable functions of the vector X such that* $P[g_t = 0] = 0$. *Then the statistic* W_g *defined in (13) is distributed as a Wilcoxon signed rank variate,* i.e., *as* $W = \sum_{t=1}^{n} tB_t$ *where* $B_1, \dots, B_n$ *are independent uniform Bernoulli variables on* $\{0,1\}$.

The distribution of *W* in Proposition (2) has been extensively tabulated by Wilcoxon, Katti and Wilcox (1970) among others, and the normal approximation with $E(W) = n(n+1)/4$ and $Var(W) = n(n+1)(2n+1)/24$ works well even for small values of *n*.

The following proposition establishes exact distribution for the statistic SR_g in (14) without the additional assumption that the vectors *Y* and *X* are independent as in Proposition 2.

Proposition 3 (Campbell and Dufour, 1995) *Let* $Y = (Y_1, \dots, Y_n)'$ *and* $X = (X_0, \dots, X_{n-1})'$ *be two* $n \times 1$ *random vectors such that (2) and (4), with* $m_0 = 0$, *hold. Let* $g_t = g_t(I_t)$, $t = 0, \dots, n-1$, *be a sequence of measurable functions of* $I_t = (X_0, X_1, \dots, X_t, Y_1, \dots, Y_t)'$ *such that* $P[g_t = 0] = 0$ *for* $t = 0, \dots, n-1$, *let*

$|Y| = (|Y_1|,\ldots,|Y_n|)'$, *and define the sign variables* $s_t = u(Y_t g_{t-1})$ *for* $t = 1,\ldots,n$. *Then the following two properties hold:*

(a) the signs $s_1,\ldots,s_n$ are mutually independent and, provided $|Y_t| \neq 0$ for $t = 1,..,n$,

$$P\left[s_t = 0 \,\middle|\, |Y|\right] = P\left[s_t = 1 \,\middle|\, |Y|\right] = 0.5; \text{for } t = 1,..,n;$$

(b) the statistic SR_g defined by (14) follows the same distribution as the Wilcoxon signed rank variate $W = \sum_{t=1}^{n} tB_t$ where $B_1,\ldots,B_n$ are independent uniform Bernoulli variables on $\{0,1\}$.

Finally, Campbell and Dufour (1995) extend the above results by relaxing totally or partially the assumptions that Y_t and X_t (or more generally g_t) have no probability mass at zero; see Campbell and Dufour (1995: Proposition 4). Their Proposition 4-(b) shows that, provided $g_0,\ldots,g_{n-1}$ have no probability mass at zero, tests based on S_g in (10) can be performed conditionally on the non-zero $Y_t's$, *i.e.* after dropping the zero $Y_t g_{t-1}$ products. For the more general case where $g_0,\ldots,g_{n-1}$ may have a mass at zero, the distribution of S_g appears difficult to determine. However, their Proposition 4-(a) shows that a simple alternative consists in replacing S_g by the closely related statistic $\bar{S}_g = \sum_{t=1}^{n} u(Y_t \bar{g}_{t-1})$, where $\bar{g}_t = g_t + \delta(g_t)$, with $\delta(x) = 1$ if $x = 0$, and $\delta(x) = 0$ if $x \neq 0$, to which the result of their Proposition 4-(b) applies. Similarly, under assumption (4) and for $m_0 = 0$, they show that one can use the statistic $\overline{SR}_g = \sum_{t=1}^{n} u(Y_t \bar{g}_{t-1}) R_{2t}^{+}$ instead of SR_g in (14); by their Proposition 4-(c), $\overline{SR}_g$ follows the usual Wilcoxon distribution.

2.2 *Sign-based Tests with Nuisance Parameters*

Campbell and Dufour (1997) extend the finite-sample nonparametric tests of Campbell and Dufour (1995) to allow for an unknown drift parameter. Their tests remain exact in the presence of general forms of feedback, non-normality and heteroskedasticity. They motivate their tests in the context of the following simple linear regression model with non-zero intercept:

$$Y_t = \mu + \beta X_{t-1} + \varepsilon_t. \tag{15}$$

They consider similar assumptions to those in Campbell and Dufour (1995), with the difference that the variable Y_t has now median m_0 instead of zero. Formally, they assume that $Y_1,\ldots,Y_n$ and $X_0,\ldots,X_{n-1}$ have continuous distributions such that:

$$Y_t \text{ is independent of } I_{t-1}, \text{ for each } t = 1,\ldots,n; \tag{16}$$

$$P[Y_t > m_0] = P[Y_t < m_0], \text{for } t = 1,\ldots n; \tag{17}$$

Assumptions (16) and (17) leave open the possibility of feedback from Y_t to current and future values of the X-variable, without specifying the form of feedback or any other property of the X-process. In addition, the variables Y_t need not be normal nor identically distributed. They also consider the following additional assumption:

$$Y_1,\ldots,Y_n \text{ have continuous distributions symmetric about } m_0. \tag{18}$$

The difference with Campbell and Dufour (1995) is the presence of an unknown median parameter m_0 which will complicate the construction of nonparametric tests. However, to obtain methods applicable for unknown m_0 Campbell and Dufour (1997) first consider the case where the nuisance parameter m_0 is known. In this case, the techniques of Campbell and Dufour (1995) can be applied to build exact nonparametric tests. Campbell and Dufour (1997) consider the sign statistic

$$S_g(m) = \sum_{t=1}^{n} u\left[(Y_t - m)\, g_{t-1}\right], \tag{19}$$

where the functions $u[.]$ and $g_{t-1}(.)$ are defined in Section 2.1. Under the further assumption in (18), they also consider an aligned signed rank statistics with general form:

$$SR_g(m) = \sum_{t=1}^{n} u\left[(Y_t - m)\, g_{t-1}\right] R_t^+(m), \tag{20}$$

where $R_t^+(m)$ is the rank of $|Y_t - m|$, *i.e.* $R_t^+(m) = \sum_{j=1}^{n} u\left(|Y_t - m| - |Y_j - m|\right)$ the rank of $|Y_t - m|$ when $|Y_1 - m|,\ldots,|Y_n - m|$ are put in ascending order. The following proposition establishes the finite-sample distributions of $S_g(m)$ in (19) and $SR_g(m)$ in (20) when $m = m_0$ is assumed to be a known parameter.

Proposition 4 (Campbell and Dufour, 1997) *Let* $Y = (Y_1,\ldots,Y_n)'$ *and* $X = (X_0,\ldots,X_{n-1})'$ *be two* $n \times 1$ *random vectors which satisfy Assumptions (16)and (17). Suppose further that* $P[Y_t = m_0] = 0$ *for* $t = 1,\ldots,n,$ *and let* $g_t = g_t(I_t),$ $t = 0,\ldots,n-1$ *be a sequence of measurable functions of* I_t *such that* $P[g_t = 0] = 0$ *for* $t = 0,\ldots,n-1.$

(a) Then the sign statistic $S_g(m_0)$ defined by (19) follows a $Bi(n,0.5)$ distribution, that is

$$P\left[S_g(m_0) = x\right] = \binom{n}{x} (1/2)^n \text{ for } x = 0,1,\ldots,n, \text{ where } \binom{n}{x} = n!/\left[x!(n-x)!\right].$$

(b) If Assumption (18) also holds, then the signed rank statistic $SR_g(m_0)$ defined in (20) is distributed as the Wilcoxon signed rank variate $W = \sum_{t=1}^{n} tB_t$, where $B_1,\ldots,B_n$ are independent Bernoulli variables such that $P[B_t = 0] = P[B_t = 1] = 1/2,$ $t = 1,\ldots,n.$

Now one has to deal with the fact that the centering parameter m_0 is unknown. To obtain provably valid finite-sample procedures for an unknown m_0, Campbell and Dufour (1997) adopt a three-stage approach introduced in Dufour (1990). First, they find an exact confidence set for the nuisance parameter m_0 which is valid at least under the null hypothesis. Second, corresponding to each value m in the confidence set, they construct a nonparametric test based on the methods discussed in Section 2.1. Third, the latter are combined with the confidence set for m_0 using Bonferroni's inequality to obtain valid nonparametric tests at the desired level α. Formally, let $CS(\alpha_1)$ be a confidence set for m_0 with level $1-\alpha_1$ ($P\left[m_0 \in CS(\alpha_1)\right] \geq 1-\alpha_1$, for $\alpha_1 < \alpha$), which is valid either on the assumption that Y_t has median m_0 for $t = 1,...,n$ or that Y_t is symmetric about m_0 for each t. Different approaches to the construction of $CS(\alpha_1)$ based on counting procedures are discussed in Campbell and Dufour (1997: 157-158). The following proposition provides probability bounds for the events that $S_g(m)$ in (19) is significant (or nonsignificant) at an appropriate level for all $m \in CS(\alpha_1)$ for both one-sided and two-sided tests, and similarly for $SR_g(m)$ in (20).

Proposition 5 (Campbell and Dufour, 1997) *Let* $Y = (Y_1,...,Y_n)'$ *and* $X = \left(X_0,...,X_{n-1}\right)'$ *be two* $n \times 1$ *random vectors which satisfy Assumptions (16) and (17) with* $P\left[Y_t = m_0\right] = 0$ *for* $t = 1,...,n,$ *and let* $g_t = g_t(I_t)$, $t = 0,...,n-1$ *be a sequence of measurable functions of* I_t *such that* $P\left[g_t = 0\right] = 0$ *for* $t = 0,...,n-1.$ *Let also* $S_g(m)$, $SR_g(m)$, $\bar{S}_g(.)$ *and* $\overline{SR}_g(.)$ *be defined by (19), (20) and*

$$P\left[S_g(m_0) > \bar{S}_g(\alpha)\right] \leq \alpha,\ P\left[SR_g(m_0) > \overline{SR}_g(\alpha)\right] \leq \alpha, \text{for any } 0 \leq \alpha \leq 1,$$

let $\tilde{S}_g(\delta) = n - \bar{S}_g(1-\delta)$ *and* $\widetilde{SR}(\delta) = \left(n(n+1)/2\right) - \overline{SR}_g(1-\delta)$ *for any* $0 \leq \delta \leq 1$, *and choose a,* a_1, a_2, *and* a_3 *in the interval* $[0,1]$ *such that* $0 \leq \alpha_2 \leq \alpha - \alpha_1 \leq \alpha + \alpha_1 \leq \alpha_3 \leq 1.$

(a) If $CS(\alpha_1)$ *is a confidence set for* m_0 *such that* $P\left[m_0 \in CS(\alpha_1)\right] \geq 1-\alpha_1$, then

$$P\left[S_g(m) > \bar{S}_g(\alpha_2), \forall m \in CS(\alpha_1)\right] \leq \alpha_1 + \alpha_2 \leq \alpha, \tag{7a}$$

$$P\left[M - S_g(m) > \bar{S}_g(\alpha_2), \forall m \in CS(\alpha_1)\right] \leq \alpha_1 + \alpha_2, \tag{7b}$$

$$P\left[\max\left\{S_g(m), M - S_g(m)\right\} > \bar{S}_g(\alpha_2/2), \forall m \in CS(\alpha_1)\right] \leq \alpha_1 + \alpha_2 \tag{7c}$$

$$P\left[S_g(m) < \tilde{S}_g(\alpha_3), \forall m \in CS(\alpha_1)\right] \leq 1 - (\alpha_3 - \alpha_1) \leq 1 - \alpha, \tag{7d}$$

$$P\left[M - S_g(m) < \tilde{S}_g(\alpha_3), \forall m \in CS(\alpha_1)\right] \leq 1 - (\alpha_3 - \alpha_1), \tag{7e}$$

$$P\left[\max\left\{S_g(m), M - S_g(m)\right\} < \tilde{S}_g(\alpha_3/2), \forall m \in CS(\alpha_1)\right] \leq 1 - (\alpha_3 - \alpha_1), \tag{7f}$$

with $M = n$.

(b) *If the additional Assumption (18) holds and* $K(\alpha_1)$ *is a confidence set for* m_0 *such that* $P\left[m_0 \in K(\alpha_1)\right] \geq 1-\alpha_1$, *then the inequalities (7a) to (7f) also hold with* $S_g(m)$ *replaced by* $SR_g(m)$, $\bar{S}_g(.)$ *by* $\overline{SR}_g(.)$, $\tilde{S}_g(.)$ *by* $\widetilde{SR}_g(.)$, $CS(\alpha_1)$ *by* $K(\alpha_1)$, *and M by* $M' = n(n21)/2$.

The above proposition suggests the following bounds test for the hypothesis that Y_t is orthogonal to past information I_{t-1}, for $t = 1,\ldots n$. Using the notations adopted in Proposition 5, Campbell and Dufour (1997) define

$$Q_L(S_g) = Inf\left\{S_g(m) : m \in CS(\alpha_1)\right\}, Q_L(SR_g) = Inf\left\{SR_g(m) : m \in K(\alpha_1)\right\} \quad (8a)$$

$$Q_U(S_g) = Sup\left\{S_g(m) : m \in CS(\alpha_1)\right\}, Q_U(SR_g) = \left\{SupSR_g(m) : m \in K(\alpha_1)\right\}. \quad (8b)$$

Using Proposition 5(a), it is clear that

$$P\left[Q_L(S_g) > \bar{S}_g(\alpha_2)\right] \le \alpha,\ P\left[Q_U(S_g) < \tilde{S}_g(\alpha_3)\right] \le 1-\alpha, \quad (8c)$$

where the conjunction of the events $Q_L(S_g) > \bar{S}_g(\alpha_2)$ and $Q_U(S_g) < \tilde{S}_g(\alpha_3)$ has probability zero, and similarly for $Q_L(SR_g)$ and $Q_U(SR_g)$. Thus, as pointed out in Campbell and Dufour (1997), a reasonable right one-sided test would reject the hypothesis of conditional independence ($H_0 : \beta = 0$) if $Q_L(S_g) > \bar{S}_g(\alpha_2)$ (alternatively, if $Q_L(SR_g) > \overline{SR}_g(\alpha_2)$), and would accept it if $Q_U(S_g) < \tilde{S}_g(\alpha_3)$ (alternatively, if $Q_U(SR_g) < \overline{SR}_g(\alpha_3)$); otherwise, the test is considered inconclusive. Based on the results of Proposition 5, Campbell and Dufour (1997) suggest to set $\alpha_2 = \alpha - \alpha_1$ and $\alpha_3 = \alpha + \alpha_1$. Now, to obtain a left one-sided test, one can proceed in exactly the same way with $S_g(m)$ replaced by $M - S_g(m) = n - S_g(m)$, and $SR_g(m)$ by $M' - SR_g(m)$. Finally, a two-sided sign test with level α is obtained by considering

$$QB_L(S_g) = Inf\left\{\max\left\{S_g(m), M - S_g(m)\right\} : m \in CS(\alpha_1)\right\},$$
$$QB_U(S_g) = Sup\left\{\max\left\{S_g(m), M - S_g(m)\right\} : m \in CS(\alpha_1)\right\},$$

and then taking $QB_L(S_g) > \bar{S}_g(\alpha_2)$, and $QB_U(S_g) < \tilde{S}_g(\alpha_3 / 2)$ as the rejection and acceptance regions, respectively.

Luger (2003) extends the nonparametric approach of Campbell and Dufour (1997) to testing for a random walk with an unknown drift. Instead of using the three-stage approach of Dufour (1990), which requires to find an exact confidence set for the nuisance parameter m_0, Luger (2003) suggests to eliminate the drift term using long differences in a way that preserves the properties of the original errors ε_t. In particular, he shows that long differencing does not introduce any correlation among the error terms as subtracting an estimated drift would.

Formally, Luger (2003) proposes a sign-based test for testing $H_0 : \beta = 1$ in the context of regression model in (15). To this end, he considers the first-difference $\Delta y_t = y_t - y_{t-1}$, for $t = 1,2...,n$. The basic building block of his testing procedure is the following quantity:

$$z_t = \Delta y_{t+l} - \Delta y_t, \text{for } t = 1,2...,l, \text{where } l = \frac{n}{2}.$$

He assumes that n is even, so that the midpoint l is an integer. As in Campbell and Dufour (1995, 1997), he considers the class of linear signed rank statistics defined by:

$$SR_l = \sum_{t=1}^{l} u\left(\Delta y_{t+l} - \Delta y_t\right) a_l\left(R_t^+\right), \tag{21}$$

where $u(.)$ is defined in the previous sections, $a_l(.)$ is some weighting function, and R_t^+ is the rank of $\left|\Delta y_{t+l} - \Delta y_t\right|$ defined in a similar way as in the previous sections.

To establish the finite-sample distribution of the test statistic in (21), Luger (2003) considers the following assumptions. He first assumes that

the density of the vector of the error terms $\varepsilon^n = (\varepsilon_1,\ldots,\varepsilon_n)'$ is symmetric. (22)

He also assumes that

$$P\left[\varepsilon^n = 0\right] = 0. \tag{23}$$

Assumptions (22) and (23) imply that the error terms may have discrete distributions provided the assumption (23) is satisfied, *i.e.*, there is no mass at zero. Furthermore, as shown in Luger (2003) several models of time-varying conditional variance, such as GARCH-type or stochastic volatility models, satisfy the multivariate symmetric assumption in (22).

Luger (2003) derives the following finite-sample distribution for the test statistic in (21) based on the following two observations: (i) under the null hypothesis ($\beta = 1$) the test statistic in (21) is a function only of $(\varepsilon_{t+l} - \varepsilon_t)$, for $t = 1,2,...,l,$ and (ii) under assumptions (22) and (23), the sign $u\left(\varepsilon_{t+l} - \varepsilon_t\right)$ is distributed as a Bernoulli variable $Bi(1,0.5)$.

Theorem 1 (Luger, 2003) *Let* $\varepsilon_1,...,\varepsilon_n$ *be a sequence of random variables that satisfy Assumptions (22) and (23). Then, the null distribution of any linear signed rank statistic defined by (21) has the property that*

$$SR_l = \sum_{t=1}^{l} u\left(\Delta y_{t+l} - \Delta y_t\right) a_l\left(R_t^+\right) \overset{d}{=} \sum_{t=1}^{l} B_i a_l(i),$$

where $B_1,...,B_l$ *are mutually independent uniform Bernoulli variables on* $\{0,1\}$.

Two special cases of the test statistic in (21) that have the usual distributions are:

$$S_l = \sum_{t=1}^{l} u\left(\Delta y_{t+l} - \Delta y_t\right), \tag{24}$$

$$W_l = \sum_{t=1}^{l} u\left(\Delta y_{t+l} - \Delta y_t\right) R_t^+, \tag{25}$$

where the first one is obtained from the score function $a_l(i) = 1$ and the second one (Wilcoxon signed rank statistic) is obtained with $a_l(i) = i.$ The following result, which is an immediate corollary to the above Theorem, provides the finite-sample distributions of the test statistics (24) and (25).

Corollary 1 (Luger, 2003) *Let the model given by (15) hold with Assumptions (22) and (23). Then, under the null* $H_0 : \beta = 1$,

(i) *The statistic* S_l *defined by (24) is distributed according to* $Bi(l, 0.5)$.

(ii) *The statistic* W_l *defined by (25) is distributed like* $W(l) = \sum_{t=1}^{l} iB_i$, *where* $B_1, \ldots, B_l$ *are mutually independent uniform Bernoulli variables on* $\{0,1\}$.

The sign-based statistics S_l and W_l have the virtues of those in Campbell and Dufour (1997): they have known finite-sample distributions, they are robust to departures from Gaussian conditions that underlie many parametric tests, and they are invariant to unknown forms of conditional heteroscedasticity. However, as pointed out by Luger (2003), the cost of these procedures is that only half the sample ($l = T/2$) is used to detect departures from the null. Using simulation experiments, Luger (2003) argues that this is still less than the cost of the Campbell and Dufour (1997) three-step approach. In other words, although the procedures proposed by Luger (2003) only use half the sample observations, their power can be considerably superior to the bounds tests of Campbell and Dufour (1997), especially for alternatives close to the null.

3. Sign-based Tests for Multiple Regression

In this section, we review several recent sign-based tests that have been proposed for testing the orthogonality between random variables in the context of linear and non-linear multiple regressions. We distinguish between tests that are valid for independent and dependent data. We start with point-optimal sign-based tests (hereafter POS test) proposed by Dufour and Taamouti (2010) for testing the parameters of linear and nonlinear multiple regression models with independent data. For these tests, we consider in turn two problems. The first one consists in testing whether the conditional median of a vector of observation is zero against a linear regression alternative. The second one tests whether the coefficients of a possibly nonlinear median regression function have a given value against another nonlinear median regression. We next discuss the sign-based tests proposed by Coudin and Dufour (2009) in the context of linear multiple regression models with dependent data.

3.1 *Sign-based Tests for Independent Data*

3.1.1 *Sign-based Tests for Testing the Zero Coefficient Hypothesis in Linear Regressions*

We consider the regression model in (1), with $f(X_t;\beta)$ is taken as a linear function of the parameters of interest:

$$Y_t = X_t'\beta + \varepsilon_t,\ t = 1, \ldots, n, \tag{26}$$

where X_t is a kx 1 vector of explanatory variables, $\beta \in \mathbb{R}^k$ is an unknown parameter vector, and the errors $\varepsilon_1,\ldots,\varepsilon_n$ are independent conditional on X with

$$P[\varepsilon_t > 0 \mid \mathbf{X}] = P[\varepsilon_t < 0 \mid \mathbf{X}] = \frac{1}{2},\ t = 1,\ldots, n, \tag{27}$$

where $\mathbf{X} = [x_1,\ldots,x_n]'$ is an $n \times k$ matrix. Assumption (27) entails that ε_t has no mass at zero, *i.e.* $P[\varepsilon_t = 0 \mid \mathbf{X}] = 0$ for all t. Suppose we wish to test the null hypothesis

$$H_0 : \beta = 0 \tag{28}$$

against the alternative hypothesis

$$H_1 : \beta = \beta_1. \tag{29}$$

Dufour and Taamouti (2010) propose the following POS test for the null hypothesis (28) against the alternative hypothesis (29). We then have the following result.

Proposition 6 (Dufour and Taamouti, 2010) *Under the assumptions (26) and (27), let H_0 and H_1 be defined by (28)-(29),*

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

where $u(.)$ *is defined in equation* (11),

$$a_t(\beta_1) = \ln\left[\frac{1 - P[\varepsilon_t \le -X_t'\beta_1 \mid \mathbf{X}]}{P[\varepsilon_t \le -X_t'\beta_1 \mid \mathbf{X}]}\right], \tag{30}$$

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

$$SL_n(\beta_1) > c_1(\beta_1) \tag{31}$$

is most powerful (conditional on X) for testing H_0 against H_1 among level-α tests based on the signs $(u(Y_1),\ldots,u(Y_n))'$.

Under H_0, the signs $u(Y_1),\ldots,u(Y_n)$ are i.i.d. according to a Bernoulli $Bi(1,0.5)$. The distribution of the test statistic only depends on the weights $a_t(\beta_1)$ and thus does not involve any nuisance parameter under the null hypothesis. In view of the nonparametric nature of assumption (27), this means that tests based on $SL_n(\beta_1)$, such as the test given by (31), are distribution-free and robust against heteroskedasticity of unknown form. It is a nonparametric *pivotal function*.

Under the alternative hypothesis, however, the power function of the test based on $SL_n(\beta_1)$ depends on the form of the distribution function of ε_t. An interesting special case is the one where $\varepsilon_1,\ldots,\varepsilon_n$ are i.i.d. according to a $N(0,1)$ distribution. Then the optimal test statistic $SL_n(\beta_1)$ takes the form:

$$SL_n^*(\beta_1) = \sum_{t=1}^{n} \ln\left[\frac{\Phi(X_t'\beta_1)}{1-\Phi(X_t'\beta_1)}\right]u(Y_t), \tag{32}$$

where $\Phi(\cdot)$ is the standard normal distribution function.

In view of the above characterization of the distribution of $SL_n(\beta_1)$, its distribution 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 (32) is a continuous variable, its quantiles are easy to compute. To simulate $SL_n^*(\beta_1)$ we first generate a sequence $\{u(y_t)\}_{i=1}^{n}$ under the null hypothesis. In particular, we generate a sequence $\{u(\varepsilon_i)\}_{i=1}^{n}$ which satisfies the condition (27). The variable $u(\varepsilon_t)$ takes only two values 0 and 1, so the computation of test statistic $SL_n^*(\beta_1)$ reduces to generating a sequence of Bernoulli random variables of given length with subsequent summation and the corresponding weights. The algorithm for implementing the POS test can be described as follows:

1. compute the test statistic $SL_n^*(\beta_1)$ based on the observed data, say $SL_n^*(\beta_1)^{(0)}$;
2. generate a sequence of Bernoulli random variables $\{u(\varepsilon_i)\}_{i=1}^{n}$ satisfying (27);
3. compute $SL_n^*(\beta_1)^{(j)}$ using $\{u(\varepsilon_i)\}_{i=1}^{n}$ and the corresponding weights $\{a_i(\beta_1)\}_{i=1}^{n}$;
4. choose B such that $\alpha(B+1)$ is an integer and repeat steps 1-3 B times;
5. compute the (1–α)–quantile, say $c(\beta_1)$, of the sequence $\left\{SL_n^*(\beta_1)^{(j)}\right\}_{j=1}^{B}$;
6. reject the null hypothesis at level α if $SL_n^*(\beta_1)^{(0)} \geq c(\beta_1)$.

3.1.2 *Sign-based tests for testing general full coefficient hypotheses in nonlinear regressions*

We now consider the nonlinear regression model in (1):

$$Y_t = f(X_t, \beta) + \varepsilon_t, t = 1,\ldots,n, \tag{33}$$

where X_t is an observable k x 1 vector of fixed explanatory variables, $f(\cdot)$ is a scalar function, $\beta \in \mathbb{R}^k$ is an unknown vector of parameters, and the errors $\varepsilon_1,\ldots,\varepsilon_n$ are independent conditional on X with a distribution that satisfies (27). Here it is not required that the parameter vector β is identified.

Suppose we wish to test the null hypothesis

$$H(\beta_0): \beta = \beta_0 \tag{34}$$

against the alternative hypothesis

$$H(\beta_1): \beta = \beta_1. \tag{35}$$

Dufour and Taamouti (2010) show that a point optimal sign-based test for $H(\beta_0)$ against $H(\beta_1)$ can be constructed as in Section 3.1.1. They derive the following sign-based test for the null hypothesis $H(\beta_0)$ against $H(\beta_1)$.

Proposition 7 (Dufour and Taamouti, 2010) *Under the assumptions (33) and (27), let* $H(\beta_0)$ *and* $H(\beta_1)$ *be defined by (34)-(35),*

$$SN_n(\beta_0 \mid \beta_1) = \sum_{t=1}^{n} \tilde{a}_t(\beta_0 \mid \beta_1) u\big(Y_t - f(X_t, \beta_0)\big), \tag{36}$$

where

$$\tilde{a}_t(\beta_0 \mid \beta_1) = \ln\left[\frac{1 - p(X_t, \beta_0, \beta_1 \mid \mathbf{X})}{p(X_t, \beta_0, \beta_1 \mid \mathbf{X})}\right],$$

and suppose the constant $c_1(\beta_0, \beta_1)$ *satisfies*

$$\mathrm{P}\left[\sum_{t=1}^{n} \tilde{a}_t(\beta_0 \mid \beta_1) u\big(Y_t - f(X_t, \beta_0)\big) > c_1(\beta_0, \beta_1)\right] = \alpha$$

under $H(\beta_0)$, with $0 < \alpha < 1$. *Then the test that rejects* $H(\beta_0)$ *when*

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

is most powerful (conditional on X) for testing $H(\beta_0)$ *against* $H(\beta_1)$ *among level-*α *tests based on the signs* $(u(Y_1 - f(X_1, \beta_0)), \ldots, u(Y_n - f(X_n, \beta_0)))'$.

The test statistic $SN_n(\beta_0 \mid \beta_1)$ in (36) depends on a particular alternative hypothesis β_1. In practice, the latter is supposed to be unknown which makes the proposed POS test unfeasible. To overcome this problem, Dufour and Taamouti (2010) propose an approach (called adaptive approach) to choose the alternative β_1 at which the power of POS test is close to the power envelope. They suggest to use what is known as " split-sample technique" to choose β_1 such that the power of POS test is close to the power envelope. The alternative hypothesis β_1 is unknown and a practical problem consists in finding its independent estimate. To make size control easier, Dufour and Taamouti (2010) estimate β_1 from a sample which is independent of the one used to compute the POS test statistic. This can be easily done by splitting the sample. The idea is to divide the sample into two independent parts and use the first one to estimate the value of the alternative and the second one to compute the POS test statistic. For more details about the above adaptive approach the reader can consult Section 4 of Dufour and Taamouti (2010).

Finally, Dufour and Taamouti (2010) have also describe how to build confidence regions with known significance level α, say $C_\beta(\alpha)$, for a vector of unknown parameters β and its individual components using the above POS tests. For more details the reader is referred to their Section 4.

3.2 *Sign-based Tests for Dependent Data*

Coudin and Dufour (2009) develop finite-sample and distribution-free sign-based tests and confidence sets for the parameters of a linear regression model, where no parametric assumption is imposed on the noise distribution. In addition to non-normality and heteroscedasticity, their set-up allows for nonlinear serial dependence of unknown forms. To build their sign tests, they first consider a *mediangale* structure – the median-based analogue of a martingale difference – under which they show that the signs of *mediangale* sequences follow a nuisance-parameter-free distribution despite the presence of non-linear dependence and heterogeneity of unknown form. The *mediangale* assumption is crucial for the construction of their tests. They distinguish between *weak* and *strict* conditional mediangale. Roughly speaking, the process of the error term $\varepsilon = \{\varepsilon_t : t = 1,2,...\}$ is a weak mediangale conditional on X if:

$$P\left[\varepsilon_1 < 0 | \mathbf{X}\right] = P\left[\varepsilon_1 > 0 | \mathbf{X}\right] \text{ and}$$
$$P\left[\varepsilon_t < 0 | \varepsilon_1,...,\varepsilon_{t-1},\mathbf{X}\right] = P\left[\varepsilon_t > 0 | \varepsilon_1,...,\varepsilon_{t-1},\mathbf{X}\right], \text{for } t > 1. \tag{37}$$

The definition of *weak* conditional mediangale allows ε_t to have a discrete distribution with a non-zero probability mass at zero. A more restrictive version, called strict conditional mediangale, imposes a zero probability mass at zero. Then, $P\left[\varepsilon_1 < 0 | \mathbf{X}\right] = P\left[\varepsilon_1 > 0 | \mathbf{X}\right] = 0.5$ and $P\left[\varepsilon_t < 0 | \varepsilon_1,...,\varepsilon_{t-1},\mathbf{X}\right] = P\left[\varepsilon_t > 0 | \varepsilon_1,...,\varepsilon_{t-1},\mathbf{X}\right] = 0.5$, for $t > 1$.

Coudin and Dufour (2009) show that for the regression model in (26) and under *strict* conditional mediangale assumption on the process ε, the residual sign vector

$$s(Y - \mathbf{X}\beta) = \left[s\left(Y_1 - X_1'\beta\right),...,s\left(Y_n - X_n'\beta\right)\right]'$$

has a nuisance-parameter-free distribution (conditional on X), *i.e.* it is a "pivotal function". This implies that its distribution is easy to simulate from a combination of n independent uniform Bernoulli variables. Consequently, any statistic of the form $T = T(s(Y - \mathbf{X}\beta), X)$ is pivotal, conditional on X. Once the form of T is specified, the distribution of the statistic T is totally determined and can also be simulated.

If we wish to test $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$, then using the above result and under H_0, $s\left(Y_t - X_t'\beta_0\right) = s\left(\varepsilon_t\right)$, $t = 1,...,n$, and conditional on X,

$$T(s(Y - \mathbf{X}\beta_0), X) \sim T\left(S_n, X\right),$$

where $S_n = \left(s_1,...,s_n\right)$ and $s_1,...,s_n \overset{i.i.d.}{\sim} \textit{Bernoulli}(1/2)$. This means that a test with level α rejects H_0 when

$$T(s(Y - \mathbf{X}\beta_0), X) > c_T(X, \alpha),$$

where $c_T(X,\alpha)$ is the $(1-\alpha)$-quantile of the distribution of $T\left(S_n, X\right)$. Coudin and Dufour (2009) extend the above result to the distributions with a positive mass at zero; see their Proposition 3.2.

As a particular case of $T(s(Y - \mathbf{X}\beta_0), X)$, they consider the following test statistic:

$$D_S\left(\beta_0, \Omega_n\right) = s(Y - \mathbf{X}\beta_0)'\mathbf{X}\Omega_n\left(s(Y - \mathbf{X}\beta_0), \mathbf{X}\right)\mathbf{X}'s(Y - \mathbf{X}\beta_0),$$

where $\Omega_n\left(s(Y - \mathbf{X}\beta_0), \mathbf{X}\right)$ is a $p \times p$ weight matrix that depends on the constrained signs $s(Y - \mathbf{X}\beta_0)$ under H_0. They argue that the weight matrix $\Omega_n\left(s(Y - \mathbf{X}\beta_0), \mathbf{X}\right)$

provides a standardization that can be useful for power considerations as well as to account for dependence schemes that cannot be eliminated by the sign transformation. Furthermore, statistics of the form $D_S(\beta_0, \Omega_n)$ include as special cases the ones studied by Koenker and Bassett (1982) and Boldin *et al.* (1997). In other words, by taking $\Omega_n = I_p$ and $\Omega_n = (\mathbf{X}'\mathbf{X})^{-1}$, we get:

$$SB(\beta_0) = s(Y - \mathbf{X}\beta_0)'\mathbf{X}\mathbf{X}'s(Y - \mathbf{X}\beta_0) = \left\|\mathbf{X}'s(Y - \mathbf{X}\beta_0)\right\|^2,$$

$$SF(\beta_0) = s(Y - \mathbf{X}\beta_0)'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'s(Y - \mathbf{X}\beta_0) = \left\|\mathbf{X}'s(Y - \mathbf{X}\beta_0)\right\|_m^2.$$

Boldin *et al.* (1997) show that $SB(\beta_0)$ and $SF(\beta_0)$ can be associated with locally most powerful tests in the case of i.i.d. disturbances under some regularity conditions on the distribution function. Coudin and Dufour (2009) have extended the proof of Boldin *et al.* (1997) to disturbances that satisfy the mediangale property and for which the conditional density at zero is the same $f_t(0|\mathbf{X}) = f(0|\mathbf{X})$ $t = 1,...,n$. They provide the following form of the locally optimal test statistic which is associated with the mean curvature, *i.e.* the test with the highest power near the null hypothesis according to a trace argument.

Proposition 8 (Coudin and Dufour, 2009) *In model (26), suppose the mediangale Assumption (37) holds, and the disturbances ε_t are heteroscedastic with conditional densities $f_t(.|\mathbf{X})$, $t = 1,2,...$, which are continuously differentiable around zero and such that $f_t'(.|\mathbf{X}) = 0$. Then, the locally optimal sign-based statistic associated with the mean curvature is*

$$\widetilde{SB}(\beta_0) = s(Y - \mathbf{X}\beta_0)'\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}'s(Y - \mathbf{X}\beta_0),$$

$$\textit{where } \widetilde{\mathbf{X}} = diag\left(f_1(.|\mathbf{X}),...,f_n(.|\mathbf{X})\right)X.$$

When $f_i(.|\mathbf{X})$'s are unknown, the optimal statistic is not feasible. In this case, the optimal weights must be replaced by approximations, such as weights derived from the normal distribution.

Coudin and Dufour (2009) discuss the implementation of the above test in the case of linearly dependent processes. In the case of discrete distribution and to reach the nominal level when using the above test, they propose to use the technique of Monte Carlo tests with a randomized tie-breaking procedure.

Finally, Coudin and Dufour (2009) discuss how to build confidence sets for the vector β or for its individual components. For more details the reader is referred to their Section 4.

4. Applications

4.1 *Testing the Long-horizon Predictability of Stock Returns*

One of the main issues of stock return predictability regressions in finance is the persistent or near-nonstationary behavior of the regressors such as dividend-price ratio, which leads to well known problems of size distortion in predictability testing, see Mankiw and Shapiro (1986). This issue has generated substantial interest in both econometrics and empirical finance; see Cavanagh, Elliott and Stock (1995); Stambaugh (1999); Campbell and Yogo (2006); Jansson and Moreira (2006), among many others.

To overcome this problem, Liu and Maynard (2007) have recently suggested to use the sign and signed rank tests of Campbell and Dufour (1995, 1997). Their motivation is that the sign-based tests provide correct size without any modeling assumptions whatsoever on the regressor. In addition, these tests offer exact finite sample inference under weak conditions.

However, Liu and Maynard (2007) point out that one practical limitation of finite sample sign and signed rank tests is that they require white noise assumptions on the dependent variable under the null hypothesis, which rules out the direct application of these robust tests to long-horizon predictability regressions. The reason that sign tests cannot be directly applied to long-horizon regressions is that the return horizon in these regressions (*e.g.* 4 years) typically exceeds the sampling frequency (*e.g.* 1 month). Thus, the returns on the left-hand side (LHS) of the predictive regression overlap for multiple periods thereby violating the required white noise assumptions.

To make the sign and signed rank tests applicable to long-horizon predictability regressions, Liu and Maynard (2007) suggest to rearrange the predictive regression considered earlier in the finance literature such as in Jegadeesh (1991) and Cochrane (1991). The latter show that the regression of a long-horizon return on a single period predictor may be replaced by a regression of a one period return on a long-horizon regressor without fundamentally altering the interpretation of the null hypothesis. Thus, replacing a long-horizon LHS variable with a long-horizon right-hand side (RHS) variable one recover the white noise assumption on the LHS variable under the null hypothesis. Formally, Liu and Maynard (2007) first consider the following traditional predictive regression:

$$Y_{t+k}^{k} = \alpha(k) + \beta(k) X_t + \varepsilon_{1,t+k}^{k}, \tag{38}$$

where $Y_{t+k}^{k} = Y_{t+1} + \ldots + Y_{t+k}$ defines the k-period return and residual, $\varepsilon_{1,t+1}^{k}$, satisfies

$$\varepsilon_{1,t+1}^{k} = \varepsilon_{1,t+1} + \ldots + \varepsilon_{1,t+k}$$

when the null hypothesis of unpredictability ($H_0 : \beta(k) = 0$) is true. Since in practice X_t is typically a persistent regressor and its process might present some correlation with the error term $\varepsilon_{1,t+1}^{k}$, this affects the statistical behavior of the OLS estimator $\hat{\beta}(k)$ and leads to invalid inference when using the classical t-test or F-test; see Cavanagh, Elliott and Stock (1995); Stambaugh (1999); Campbell and Yogo (2006); Jansson and Moreira (2006) among many others.

Because of the persistence in X_t and the loss of white noise assumption on the dependent variable Y_{t+k}^{k} under the null hypothesis, instead of employing the sign and signed rank methods to test (38) directly, Liu and Maynard (2007) instead follow an approach similar to that of Jegadeesh (1991) and Cochrane (1991) who base their test of $\beta(k)=0$ on a simple rearrangement of (38) under the null hypothesis, that avoids the serial correlation in the residuals. They define a long-horizon version of the regressor X_t as:

$$X_t^k = X_{t-k+1} + X_{t-k+2} + \ldots + X_t$$

and they show that when X_t is stationary, the long-horizon non-predictability restriction $\beta(k)=0$ is equivalent to the orthogonality condition $cov(Y_{t+k}^k, X_t)=0$ and

$$cov(Y_{t+k}^k, X_t) = cov(Y_{t+1}, X_t^k),$$

where the latter covariance is the numerator of the slope coefficient $\gamma(k)$ in the regression of Y_{t+1} on X_t^k:

$$Y_{t+1} = \gamma_0(k) + \gamma(k) X_t^k + v_{t+1}. \tag{39}$$

Thus, the restriction of the null hypothesis, $\beta(k)=0$ in (38) is equivalent to the null restriction that $\gamma(k)=0$ in (39). Consequently, Liu and Maynard (2007) test the null hypothesis $\beta(k)=0$ using the following sign and signed rank test statistics:

$$S_n^k = \sum_{t=1}^{n-1} u\left[\left(Y_{t+1} - m_0\right) X_t^{k*}\right],\ SR_n^k = \sum_{t=1}^{n-1} u\left[\left(Y_{t+1} - m_0\right) X_t^{k*}\right] R_{t+1}^{+}\left(m_0\right),$$

where the functions $u[.]$ and R_{t+1}^{+} are defined in the previous sections, m_0 is the unconditional median for Y_t, and $X_t^{k*} \equiv X_t^k - med_1^t\left(X_t^k\right)$ is the value of X_t^k centred about the sample median of $X_1^k, \ldots, X_t^k$. Campbell and Dufour (1997) argue that centering of this type is known to improve test power, but does not affect size as $med_1^t\left(X_t^k\right)$ is predetermined. The finite sample distributions of S_n^k and SR_n^k which one can use to make a decision about H_0 are defined in Campbell and Dufour (1997).

Using the one-month treasury bill and the dividend-price ratio as predictors of stock returns, with return horizons ranging from one-month to four years, Liu and Maynard (2007) confirm the existing evidence of stock return predictability using the treasury bill at short to medium horizons, but find no significant evidence of predictability at either short or long-horizons employing the dividend-price ratio as a predictor.

4.2 *Testing the Mean-variance Efficiency*

Using the results of Luger (2003) discussed above, Gungor and Luger (2009) develop exact distribution-free sign-based tests of unconditional mean-variance efficiency. To derive their tests, Gungor and Luger (2009) consider the following traditionally used excess-return system of equations:

$$r_{it} = \alpha_i + \beta_i r_{pt} + \varepsilon_{it}, \text{ for } t = 1, \ldots, T \text{ and } i = 1, \ldots, N, \tag{40}$$

where r_{it} and r_{pt} are the time-t returns on asset i and portfolio p, respectively, in excess of the riskless rate, and ε_{it} is a random error term for asset i in period t with the property that $E\left[\varepsilon_{it}\right]=0$.

The mean-variance efficiency condition that states that

$$E\left[r_{it}\right]=\beta_i r_{pt},\ i=1,\ldots,N, \tag{41}$$

can be assessed by testing:

$$H_0:\alpha_i=0,\ i=1,\ldots,N, \tag{42}$$

in the regression equation (40). This null hypothesis follows from a comparison of the unconditional expectation in (41) to the mean-variance efficiency condition in equation (40). If H_0 does not hold, it would be possible to obtain a higher expected return with no higher risk, contradicting the hypothesis that portfolio p is mean-variance efficient.

To test H_0 using exact sign-based tests, Gungor and Luger (2009) first consider the following transformation of the regression model in (40)

$$\frac{r_{it}}{r_{pt}}=\frac{\alpha_i}{r_{pt}}+\beta_i+\frac{\varepsilon_{it}}{r_{pt}},$$

where the slope parameter β_i is viewed now as an intercept. Thereafter, the nuisance parameter βi can be eliminated from the inference problem via the long differences, and this leads to the following new regression:

$$d_t^{\left(\frac{T}{2}\right)}=\alpha_i x_{pt}+\left(\frac{\varepsilon_{it+\frac{T}{2}}}{r_{p,t+\frac{T}{2}}}-\frac{\varepsilon_{it}}{r_{pt}}\right), \tag{43}$$

where the new dependent variable $d_t^{\left(\frac{T}{2}\right)}=r_{i,t+\frac{T}{2}}/r_{p,t+\frac{T}{2}}-r_{it}/r_{pt}$ and the new independent variable $x_{pt}=\left(r_{pt}-r_{p,t+\frac{T}{2}}\right)/\ r_{pt}r_{p,t+\frac{T}{2}}$. Using the results in Luger (2003) and based on the regression model in, Gungor and Luger (2009) suggest to test H_0 using the test statistics

$$SB=\max_{1\le i\le N}\left|S_i\right|,\ WB=\max_{1\le i\le N}\left|SR_i\right|,$$

where the sign-based statistics $S_i=\sum_{t=1}^{T/2}u\left(d_t^{\left(\frac{T}{2}\right)}x_{pt}\right)$ and $SR_i=\sum_{t=1}^{T/2}u\left(d_t^{\left(\frac{T}{2}\right)}x_{pt}\right)R_t^{+}$. They also consider the following asymptotic versions of the test statistics SB and WB that are based on the normally distributed approximations of the statistics S_i and SR_i:

$$SB^*=\max_{1\le i\le N}\left|S_i^*\right|,\ WB^*=\max_{1\le i\le N}\left|SR_i^*\right|, \tag{44}$$

where

$$S_i^* = \frac{S_i - T/4}{\sqrt{T/8}}, \quad SR_i^* = \frac{SR_i - T(T+2)/16}{\sqrt{T(T+2)(T+1)/96}}. \tag{45}$$

As shown in Gungor and Luger (2009), the maximal statistics in (44) correspond to the ones with the smallest p-values, since the individual test statistics in (45) are identically distributed. The motivation behind using the maximal statistics is because H_0 in (42) can be viewed as the intersection of the N subhypotheses $H_{0i}: \alpha_i = 0,\ i = 1,...,N$. Consequently, the decision rule is then built from the equivalence that H_0 is false if any of its subhypotheses is false; *i.e.*, one rejects H_0 if any one of the separate tests, say $S_1^*,\ldots, S_N^*$, rejects it.

Finally, Gungor and Luger (2009), based on the results of Sidak (1967), argue that the asymptotic marginal null distributions of SB^* and WB^* satisfy the inequalities

$$P\left[SB^* \leq \omega_{\alpha^*/2}\right] \geq (1-\alpha) \text{ and } P\left[WB^* \leq \omega_{\alpha^*/2}\right] \geq (1-\alpha), \tag{46}$$

where $\omega_{\alpha^*/2}$ is the upper $\alpha^*/2$ critical point of the standard normal distribution and $\alpha^* = 1-(1-\alpha)^{1/N}$. The above inequalities in (46) indicate that asymptotically the level of the test of H_0 that compares either SB^* or WB^* to $\omega_{\alpha^*/2}$ is equal to α. This means that if the ordinary two-sided p-value of SB^* or WB^* is, say pv, then the multiplicity-adjusted two-sided p-value is calculated from the equation $pv^* = 1-(1-pv)^N$.

Finally, an extension of the mean-variance efficiency sign-based test of Gungor and Luger (2009) can be found in Gungor and Luger (2013). The latter provide a sign-based statistical procedure that allows one to test the beta-pricing representation of linear factor pricing models, instead of the single market factor model in (40). Exploiting results from Coudin and Dufour (2009), Gungor and Luger (2013) obtain tests of multi-beta pricing representations that relax three assumptions of the prominent mean-variance efficiency test of Gibbons, Ross, and Shanken (1989): (i) the assumption of identically distributed disturbances, (ii) the assumption of normally distributed disturbances, and (iii) the restriction on the number of assets. A very attractive feature of Gungor and Luger's (2013) test is that it is applicable even if the number of assets is greater than the length of the time series. This stands in sharp contrast to the Gibbons, Ross, and Shanken's (1989) test and other approaches that are based on usual estimates of the disturbance covariance matrix. It is worth mentioning that, the main drawback of Gibbons, Ross, and Shanken's (1989) approach is that to avoid singularities and be computable, this test requires the size of the cross section (number of assets) to be less than that of the time series. Consequently, the power of this test and others is negatively affected by the number of assets under consideration. In other words, the number of covariances that need to be estimated grows rapidly with the number of included assets. As a result, the precision with which this increasing number of parameters can be estimated deteriorates given a fixed time-series length, which decreases the power of the tests. In contrast, a simulation experiment that compares the performance of the Gungor and Luger's (2013) test with several standard tests, including Gibbons, Ross, and Shanken's (1989) test, shows that the power of Gungor and Luger's (2013) test increases as the cross section becomes larger.

Conclusion

We have reviewed several finite-sample sign-based tests for testing the orthogonality between random variables in the context of linear and nonlinear regression models. The sign tests are very useful when the data at the hands contain few observations, are robust against heteroskedasticity of unknown form, and can be used in the presence of non-Gaussian errors. These tests are also flexible since they do not require the existence of moments for the dependent variable and there is no need to specify the nature of the feedback between the dependent variable and the current and future values of the independent variable. Finally, we discussed several applications where the sign-based tests can be used to test for multi-horizon predictability of stock returns and for the market efficiency.

REFERENCES

Bahadur, R., and L. J. Savage (1956). "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems." *Annals of Mathematical Statistics,* 27: 1115-1122.

Boldin, M. V., G. I. Simonova, and Y. N. Tyurin (1997). "Sign-based Methods in Linear Statistical Models." in *Translations of Mathematical Monographs*, Volume 162. Maryland: American Mathematical Society.."

Campbell, B., and J. M. Dufour (1991). "Overrejections in Rational Expectations Models: A Non-parametric Approach to the Mankiw-Shapiro Problem." *Economic Letters,* 35: 285-290.

Campbell, B., and J. M. Dufour (1995). "Exact Nonparametric Orthogonality and Random Walk Tests." *Review of Economics and Statistics,* 77: 1-16.

Campbell, B., and J. M. Dufour (1997). "Exact Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter." *International Economic Review,* 38: 151-173.

Campbell, J., and M. Yogo (2006). "Efficient Tests of Stock Return Predictability." *Journal of Financial Economics,* 81: 27-60.

Cavanagh, C. L., G. Elliott, and J. H. Stock (1995). "Inference in Models with Nearly Integrated Regressors." *Econometric Theory,* 11: 1131-1147.

Cochrane, J. (1991). "Volatility Tests and Efficient Markets: A Review Essay." *Journal of Monetary Economics,* 27: 463-487.

Coudin, E., and J. M. Dufour (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.-M. (1990). "Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors." *Econometrica, 58*: 475-494.

DUFOUR, J.-M. (2003). "Identification, Weak Instruments and Statistical Inference in Econometrics." *Canadian Journal of Economics,* 36: 767-808.

DUFOUR, J.-M., and M. HALLIN (1991). "Nonuniform Bounds for Nonparametric Tests." *Econometric Theory,* 7: 253-263.

DUFOUR, J.-M., F. JOUNEAU, and O. TORRÈS (2008). "On (non)-testability: Applications to Linear and NonlinearSemiparametric and Nonparametric Regression Models." Technical report, Department of Economics and CIREQ, McGill University, and GREMARS, Université de Lillie 3.

DUFOUR, J. M., and A. TAAMOUTI (2010). "Exact Optimal Inference in Regression Models under Heteroskedasticity and Non-normality of Unknown Form." *Computational Statistics & Data Analysis*, 54: 2532-2553.

GIBBONS, M. R., S. A. ROSS, and J. SHANKEN (1989). "A Test of the Efficiency of a Given Portfolio." *Econometrica*, 57: 1121-1152.

GUNGOR, S., and R. LUGER (2009). "Exact Distribution-free Tests of Mean-variance Efficiency." *Journal of Empirical Finance*, 16: 816-829.

GUNGOR, S., and R. LUGER (2013). "Testing Linear Factor Pricing Models with Large Cross Sections: A Distribution-free Approach." *Journal of Business and Economic Statistics*, 31: 66-77.

JANSSON, M., and M. J. MOREIRA (2006). "Optimal Inference in Regression Models with Nearly Integrated Regressors." *Econometrica*, 74: 681-715.

JEGADEESH, N. (1991). "Seasonality in Stock Price Mean Reversion: Evidence from the U.S., and the U.K." *Journal of Finance,* 46: 1427-1444.

KOENKER, R., and G. BASSETT, Jr. (1982). "Tests of Linear Hypotheses and L1 Estimation." *Econometrica,* 50: 1577-84.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. New York: John Wiley.

LEHMANN, E., and C. STEIN (1949). "On the Theory of some Non-parametric Hypotheses." *Annals of Mathematical Statistics,* 20: 28-45.

LIU, W., and A. S. MAYNARD (2007). " A New Application of Exact Nonparametric Methods to Long-horizon Predictability Tests." *Studies in Nonlinear Dynamics & Econometrics*, 11: 163-199.

LUGER, R. (2003). "Exact Non-parametric Tests for a Random Walk with Unknown Drift under Conditional Heteroscedasticity." *Journal of Econometrics*, 115: 259-276.

MANKIW, N. G., and M. SHAPIRO (1986). "Do we Reject too often? Small Sample Properties of Tests of Rational Expectations Models." *Economic Letters,* 20: 139-145.

PRATT, J., and J. GIBBONS (1981). *Concepts of Nonparametric Theory.* New York: Springer-Verlag.

ROUSSEEUW, P. J., and A. LEROY (1987). *Robust Regression and Outlier Detection*, Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

ROUSSEEUW, P. J., and V. YOHAI (1984). "Robust Regression by Means of S-Estimators." in *Robust and Nonlinear Time Series Analysis*, edited by J. FRANKE, W. HARDLE, and D. MARTIN, Lecture Notes in Statistics No. 26, Springer Verlag, Berlin.

SIDAK, Z. (1967). "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions." *Journal of the American Statistical Association,* 62: 626-633.

STAMBAUGH, R. (1999). "Predictive Regressions." *Journal of Financial Economics,* 54: 375-421.

YOHAI, V. J., and R. ZAMAR (1988). "High Breakdown Point Estimates of Regression by Means of the Minimization of an Efficient Scale." *Journal of the American Statistical Association,* 83: 406-413.

WHITE, H. (1980). "A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica,* 48: 817-838.