WORKING PAPERS

26 | 2011

A CLASS OF ROBUST TESTS IN AUGMENTED PREDICTIVE REGRESSIONS

Paulo M.M. Rodrigues

Antonio Rubia

October 2011

The analyses, opinions and findings of these papers represent the views of the authors, they are not necessarily those of the Banco de Portugal or the Eurosystem

Please address correspondence to
Paulo M.M. Rodrigues
Banco de Portugal, Economics and Research Department
Av. Almirante Reis 71, 1150-012 Lisboa, Portugal;
Tel.: 351 21 313 0831, email: Paulo M.M. Rodrigues@bportugal.pt

Banco de Portugal
EUROSYSTEM

BANCO DE PORTUGAL
Av. Almirante Reis, 71
1150-012 Lisboa
www.bportugal.pt

Edition
Economics and Research Department

Pre-press and Distribution
Administrative Services Department
Documentation, Editing and Museum Division
Editing and Publishing Unit

Printing
Administrative Services Department
Logistics Division

Lisbon, October 2011

Number of copies
80

ISBN 978-989-678-106-4
ISSN 0870-0117 (print)
ISSN 2182-0422 (online)
Legal Deposit no. 3664/83

A Class of Robust Tests in Augmented Predictive Regressions[*]

Paulo M.M. Rodrigues[†] Antonio Rubia[‡]

Abstract

This paper focuses on the analytical discussion of a robust t-test for predictability and on the analysis of its finite-sample properties. Our analysis shows that the procedure proposed exhibits approximately correct size even in fairly small samples. Furthermore, the test is well-behaved under short-run dependence, and can exhibit improved power performance over alternative procedures. These appealing properties, together with the fact that the test can be applied in a simple and direct way in the linear regression context, suggests that the modified t-statistic introduced in this paper is well suited for addressing predictability in empirical applications.

Keywords: Predictability, persistence, contemporaneous correlation, stock returns.
JEL: C12, C22.

---

[*] **Acknowledgements:** The authors thank the participants of the XVII Foro de Finanzas held in Elche, Spain, the JSM 2010 meeting held in Vancouver, Canada, the 3rd European Time Series Network workshop held at the University of Nottingham, the 3rd Portuguese Economic Journal conference held at the University of Madeira and seminar participants at the Department of Mathematics of Instituto Superior Técnico, as well as Juan Mora, Francisco Peñaranda, Enrique Sentana for useful comments and suggestions. Any remaining errors are our own responsability. Financial support from POCTI/ FEDER (grant ref. PTDC/ECO/64595/2006) and the ECO2008-02599/ECON projects is gratefully acknowledged.

[†] Banco de Portugal, Economics and Research Department; NOVA School of Business and Economics, Universidade Nova de Lisboa and CEFAGE. **E-mail:** pmrodrigues@bportugal.pt

[‡] Department of Financial Economics, University of Alicante. **E-mail:** antonio.rubia@ua.es

1 Introduction

Predictive regressions are widely used in finance to address the existence of time-varying predictable patterns. In this analysis, the main variable of interest, say monthly excess of market returns, is usually regressed on lagged values of a posited predictor, such as the log dividend yield, and predictability formally judged through the significance of the resultant estimates; see Campbell (2008) for an overview. However, in practice two statistical considerations make this analysis difficult: i) predictors typically exhibit highly persistent dynamics; and ii) predictors' innovations are typically largely correlated with the innovations of the variable to be predicted, which raises a problem of simultaneity in regression. In small samples, this feature may lead to sizeable biases in the least-squares estimates of the predictive parameter (Stambaugh, 1986, Mankiw and Shapiro, 1986) and substantial overrejections in the corresponding analysis of significance (Elliott and Stock, 1994, Cavanagh, Elliot and Stock, 1995). A number of alternative approaches to ensure valid inference have been suggested in the literature, among which the class of augmented predictive regressions which is of major interest for the present paper. The different tests in this category can be related to the estimates of a predictive regression that is augmented with an empirical estimate of the innovations to the predictor, and differ mainly in their basic assumptions about the stochastic properties of the predictor. For instance, the procedures in Lewellen (2004), Amihud and Hurvich (2004), and Amihud, Hurvich, and Wang (2010) assume stationarity, while the test proposed by Campbell and Yogo (2006) is derived under near-integration. These assumptions have sharp implications on the tests and their theoretical properties.

In this paper, we provide a detailed discussion, of the theoretical properties of the test statistics related to augmented regressions, extending existing literature and clarifying some important issues. Furthermore, we propose a modified t-statistic for augmented (and non-augmented) regressions which renders valid inference in the empirical context that characterizes predictability analysis. The distinctive feature of our test is that it ensures, by construction, inference with approximately correct nominal size at any arbitrary significance level and independently of the stochastic properties of the predictive variable, *i.e.,* whether the predictor is stationary, near-integrated, or integrated. The test displays robustness to the formal uncertainty of the stochastic properties of the predictor which we refer to as quantile invariance in this paper. This is achieved by scaling the standard t-statistic with a stochastic factor that endogenously smoothens the discontinuities in the limit distributions as the order of integration is allowed to go from stationarity to integrated, following a strategy suggested by Vogelsang (1998) in the different context of unit root testing; see also Harvey, Leybourne and Taylor (2006). In our context, this approach results in test statistics whose null critical values ensure correct size properties by construction, and which can be applied on predictors that show different degrees of persistence. In contrast, available procedures in the augmented-regression literature are intended for either strongly persistent or stationary processes and may not control size adequately otherwise. Monte Carlo analysis reveals that our procedure ensures approximately correct size even in small samples under different types of data generating processes, and exhibits good power performance, both in absolute terms and in relation to other alternative tests. These appealing properties suggest that the modified t-statistic introduced in this paper is well suited for addressing predictability in empirical applications.

The importance of robustness in predictive analysis deserves to be carefully discussed.

Assuming stationarity makes direct inference possible, which, nevertheless, may not be exempt of important size distortions when the predictors are strongly persistent. On the other hand, allowing for more general near-integrated patterns requires the use of other procedures necessary to make inference feasible in practice, such as, for instance, the Bonferroni-type confidence intervals considered in Campbell and Yogo (2006). Thus, the finite-sample size properties of these tests are strongly conditioned by the true dynamics of the predictor. In contrast, our test ensures approximately correct empirical size irrespectively of whether the predictor is stationary or not. To the best of our knowledge, these properties are only shared by the nonparametric test recently proposed by Maynard and Shimotsu (2009), although the methodological approach differs considerably. Our test has the comparative advantage of building directly on the simplest context of linear regressions and of being as easy to implement as the conventional t-test, exhibiting better power properties in the context that usually describes predictive regressions. On the other hand, the test in Maynard and Shimotsu (2009) will hold in more general contexts than those described here.

The rest of the paper is organized as follows. Section 2 introduces the predictive regression framework considered in this paper and discusses general properties in the augmented context. Section 3 introduces the robustified t-test and discusses its large-sample distribution under the null and alternative hypotheses. Section 4 presents Monte Carlo results on the size and power performance of the new test as well as a comparison with available procedures. Finite sample critical values are also presented in this section. Section 5 contains the results of an empirical application. Section 6 summarizes and concludes. Finally, an appendix collects the proofs of the results put forward throughout the paper.

In the sequel, '$\Rightarrow$' and '$\xrightarrow{p}$' are used to denote weak convergence of the associated probability measures and convergence in probability, respectively, as the sample size is allowed to grow unbounded. The conventional notation $o(1)$ ($o_p(1)$) is used to represent a series of numbers (random numbers) converging to zero (in probability). Similarly, $O(1)$ ($O_p(1)$) denotes a series of numbers (random numbers) that are bounded (in probability).

2 Inference in predictive regressions

2.1 Basic assumptions and notation

Predictive regressions are used to determine whether an observable time series, say $\{y_t\}$, can be linearly predicted using lagged values of a conditioning variable, $\{x_t\}$. This, basically consists of testing the null hypothesis, $\mathrm{H}_0 : \beta = 0$, in the regression

$$y_t = \alpha + \beta x_{t-1} + u_t, \qquad t = 1, ..., T \tag{1}$$

with u_t denoting a disturbance term; see, for instance, Fama and French (1988) and Campbell and Shiller (1988).

As previously discussed, certain empirical features that characterize predictive regressions can complicate statistical inference considerably and, therefore, should be acknowledged explicitly in the theoretical analysis. Firstly, most predictors typically considered in the literature exhibit strongly-persistent dynamics to the extent that autoregressive unit roots are often difficult to reject. Since the seminal work of Stambaugh (1999), theoretical studies have mostly accommodated this empirical feature parsimoniously by assuming first-order autoregressive dynamics in $\{x_t\}$, *viz.*,

$$x_t = \mu + \phi x_{t-1} + v_t, \quad t = 1, ..., T \tag{2}$$

considering different sets of restrictions that characterize ϕ and the innovation vector $\xi_t = (u_t, v_t)'$. Secondly, the sample covariance of $\widehat{\xi}_t$ often reveals that the shocks to the system are highly contemporaneously correlated.[1] The following assumptions lay out all of these features and shall be maintained throughout the following subsections. We shall consider more general restrictions later in Section 3.

Assumption 1. *The data generating process of* $\{y_t\}$ *and its posited predictor* $\{x_t\}$ *is formed by equations* (1)-(2), *with* x_s *random or constant for all* $s \leq 0$.
Assumption 2. *The elements of the innovation vector* $\xi_t = (u_t, v_t)'$ *in the two-equation system are serially and independently Gaussian distributed with* $E(\xi_t) = 0$, $E(\xi_t \xi_t') = \Sigma$,

$$\Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}$$

and $\sigma_{uv} = \rho\sigma_u\sigma_v$ *for some correlation parameter* $\rho \in (-1, 1)$.
Assumption 3. *(Stationarity) The autoregressive dynamics of the predictor is characterized by a fixed coefficient* $|\phi| < 1$.
Assumption 3'. *(Near-Integration) The autoregressive dynamics of the predictor is characterized by the coefficient* $\phi = 1 + c/T$, *for some constant value* $c \leq 0$.

Assumptions 1-3 characterize the theoretical setting analyzed in Stambaugh (1999), which has been widely used in subsequent literature. Although Stambaugh's setting relies on highly restrictive assumptions, it captures the basic issues of predictive regressions and facilitates its theoretical discussion. However, some of these restrictions are not essential for the results and shall be generalized considerably in Section 3. The main purpose of Assumptions 3 and 3' is to characterize the long-run properties of the predictor in two different settings: stationarity and near-integration. Stationarity simplifies the theoretical discussion considerably, as it allows to invoke standard first-order asymptotic distribution theory arguments. It defines the correct setting for mean-reverting predictors but may result in large-sample representations that do not work well in small samples when the largest autoregressive root of $\{x_t\}$ is in the neighborhood of unity. In contrast, local-to-unity asymptotic theory provides a more accurate approach to the small-sample distribution of the relevant test statistics in this context. This setting is characterized by a nuisance term, the so-called local-to-unity parameter, c, which captures deviations from the unit root in a decreasing neighborhood as T increases, thereby determining the extent of serial dependence in variables that although highly persistent do not necessarily have an exact unit root; see Phillips (1987) for technical details. The unit-root case is nested as a particular case $(c = 0)$, whereas stationarity arises when c is allowed to diverge to minus infinity. The price for this theoretical flexibility, however, is that the empirical analysis becomes more difficult because c cannot be consistently estimated.[2]

---

[1] The innovations in predictive variables defined as the ratio of accounting measures of cash flow to market valuations variables typically exhibit strong negative correlation with innovations in returns. This empirical feature does not always hold. For instance, innovations in interest rates are nearly uncorrelated with returns; see Campbell (2008).

[2] Although it would be tempting to estimate c as $T(\widehat{\rho} - 1)$, Philllips (1987) shows that the resulting

2.2 Augmented predictive regressions

Stambaugh (1999) shows that the exact OLS bias of β in (1), under Assumptions 1-3, is $E\left(\widehat{\beta}-\beta\right)=\beta_{uv}E\left(\widehat{\phi}-\phi\right)$, with $\beta_{uv}\equiv\sigma_{uv}/\sigma_v^2$, and $\widehat{\phi}$ denoting the OLS estimate of ϕ in (2). Since $\widehat{\phi}$ is downward biased in small-samples and the innovations u_t and v_t are highly correlated, the autoregressive parameter bias feeds into the small-sample distribution of $\widehat{\beta}$. To correct, or at least to mitigate, this effect, Amihud and Hurvich (2004) propose a simple statistical devise that builds upon the OLS estimates from an augmented version of (1); see also Lewellen (2004) and Campbell and Yogo (2006).

To briefly introduce the basic testing strategy that underlies augmented predictive regressions, note that under Assumption 2 the orthogonal decomposition allows to write $u_t=\beta_{uv}v_t+\varepsilon_t$, with ε_t being a random error term orthogonal to v_t and independent of x_t. Thus, using this property, equation (1) may be rewritten as

$$y_t=\alpha+\beta x_{t-1}+\beta_{uv}v_t+\varepsilon_t \tag{3}$$

so that $y_t=E\left(y_t|x_{t-1}\right)+E\left(u_t|v_t\right)+\varepsilon_t$, with β still capturing the extent of *ex ante* predictability in the system. The importance of this regression results from the fact that the resultant estimates have correct statistical properties, *i.e.*, under Assumptions 1-3, Amihud and Hurvich (2004) show that the infeasible estimator of $(\beta,\beta_{uv})'$ is exactly unbiased. However, since the estimator of (3) is infeasible, Amihud and Hurvich (2004) suggest the feasible counterpart,

$$y_t=\alpha+\beta x_{t-1}+\beta_{uv}\widehat{v}_t+\varepsilon_t \tag{4}$$

with $\widehat{v}_t=x_t-\widehat{\mu}-\widehat{\phi}x_{t-1}$ denoting the OLS residuals from (2).[3]

In this paper, we consider a simple and direct variation of equations (3) and its feasible counterpart (4). This approach delivers estimates with the same statistical properties, both in small and large samples, but allows us to considerably simplify the notation and economize the mathematical derivation in the formal proofs. In particular, starting from the infeasible representation (3) we can write,

$$\left[y_t-E\left(u_t|v_t\right)\right]=E\left(y_t|x_{t-1}\right)+\varepsilon_t, \tag{5}$$

and then consider its feasible counterpart,

$$y_t^*=\alpha+\beta x_{t-1}+\varepsilon_t \tag{6}$$

where,

$$y_t^*=y_t-\widehat{\beta}_{uv}\widehat{v}_t \tag{7}$$

with $\widehat{\beta}_{uv}$ denoting the estimate of β_{uv} in the OLS regression of $\widehat{u}_t$ on $\widehat{v}_t$.

This alternative representation provides a convenient way to think about augmented predictive regressions. As discussed in Campbell and Yogo (2006), the unobservable

---

value behaves as a random variable rather than a fixed term, *i.e.*, the estimate is inconsistent. This is the natural consequence of econometrically acknowledging the uncertainty about the stochastic nature of the predictor: even if we had a sample arbitrarily large, we would not be sure about whether the predictor is really stationary or integrated.

[3] In particular, $\widehat{v}_t$ is determined after estimating $\widehat{\phi}$ with a small-sample bias correction in the spirit of Stambaugh (1999), and the standard errors of the resultant estimate of β computed according to a low-bias finite sample approximation.

process $y_t - E(u_t|v_t)$ results from subtracting off the part of the innovation to the predictor that is correlated with y_t. This yields a less noisy dependent variable and, therefore, leads to power advantages over conventional predictive analysis that stem from a relative gain in statistical efficiency. Note that, since $E(\varepsilon_t^2) = (1-\rho^2)\,\sigma_u^2$, the larger the degree of endogenous correlation in the system, the larger the amount of variability in the regressand not related to the lagged value of x_t (*i.e.*, noise variability) that can be filtered out – conversely, we can think of the standard predictive regression analysis as a particularly inefficient tool to detect predictability when ρ is large in absolute value. Since $[y_t - E(u_t|v_t)]$ cannot be directly observed, the feasible representation uses the OLS-based proxy y_t^* in (7).

The feasible estimator of β in (6) is

$$\widehat{\beta}_F = \frac{\sum_{t=2}^{T} \tilde{y}_t^* \tilde{x}_{t-1}}{\sum_{t=2}^{T} \tilde{x}_{t-1}^2} \tag{8}$$

with $\tilde{x}_t = x_t - \bar{x}$, $\bar{x} = T^{-1}\sum_{t=1}^{T} x_t$, denoting the demeaned process, and $\tilde{y}_t^* = y_t^* - \bar{y}^*$ defined analogously. Our main interest lies in the statistical properties of the OLS t-statistic for the null hypothesis $\mathrm{H}_0 : \beta = 0$ obtained from this regression, namely,

$$t_{\widehat{\beta}_F} = \frac{\sum_{t=2}^{T} \tilde{x}_{t-1}\tilde{y}_t^*}{\widehat{\sigma}_\varepsilon \left(\sum_{t=2}^{T} \tilde{x}_{t-1}^2\right)^{1/2}} = \frac{\sum_{t=2}^{T} \tilde{x}_{t-1}\left[\tilde{y}_t - \widehat{\beta}_{uv}\widehat{v}_t\right]}{\widehat{\sigma}_\varepsilon \left(\sum_{t=2}^{T} \tilde{x}_{t-1}^2\right)^{1/2}} \tag{9}$$

where $\widehat{\sigma}_\varepsilon^2$ denotes the OLS variance estimate of the regression residuals.

Remark 2.1. Augmentation of the standard regression analysis to improve the power performance of the tests has been proposed in other areas of econometrics. For instance, Hansen (1995) and Caporale and Pittis (1999) suggest using covariates to augment the Dickey-Fuller regression for unit-root testing. The critical difference in that context is that the covariates are directly observable. As we shall remark in the theoretical discussion that follows, replacing v_t by $\widehat{v}_t$ (or the process $y_t - E(u_t|v_t)$ by y_t^*) has important implications on the statistical properties, even asymptotically. Rodrigues and Rubia (2011) point out that under stationary conditions, the difference in the mean square error of the infeasible and feasible estimates, $MSE\left(\widehat{\beta}_I\right) - MSE\left(\widehat{\beta}_F\right)$, equals $\beta_{uv}^2 MSE\left(\widehat{\phi}\right)$. Thus, the use of bias-reduction techniques in the estimation of the first-order autocorrelation parameter ϕ and the residuals $\widehat{v}_t$ in order to possibly improve the small-sample properties of $\widehat{\beta}_F$ is recommended.[4]

2.3 Asymptotic properties in augmented regression testing

In this subsection, we provide the following Theorem that presents the limit results of the t-statistic computed from the infeasible and feasible augmented regressions. In this Theorem, $W(r)$ denotes a standard Wiener process, $r \in [0,1]$; $J_c(r)$ and $\overline{J}_c(r) \equiv J_c(r)-\int_0^1 J_c(s)ds$ denote, respectively, an Ornstein-Uhlenbeck and a 'demeaned' Ornstein-Uhlenbeck diffusion process on the space $D[0,1]$ of cadlag functions, with the former being characterized by the stochastic differential equation $dJ_c(r) = cJ_c(r)dr + dW(r)$, with $J_c(0) = 0$; finally $\mathcal{Z}$ is a random variable with standard normal distribution independent of $W(r)$.

---

[4] If the small-sample bias in $\widehat{v}_t$ is reduced through a suitable choice of the correcting factor ω, this feeds into the resultant estimate of the slope predictive coefficient. For instance, under the normality assumption, the autoregressive estimate is biased downward by roughly $-(1+3\phi)/T$, so Amihud and Hurvich (2004) suggest $\omega = \left(1+3\widehat{\phi}\right)/T$, or the higher-order refinement $\omega = \left(1+3\widehat{\phi}\right)\left[T^{-1}+3T^{-2}\right]$ to remove the bias in $\widehat{\beta}_F$.

Theorem 2.1. *Let $t_{\widehat{\beta}_I}$ and $t_{\widehat{\beta}_F}$ denote the least-squares t-statistics for $H_0 : \beta = 0$ computed from the infeasible and feasible augmented predictive regressions* (3) *and* (6), *respectively. Under Assumptions 1-2, and either 3 or 3', as $T \to \infty$,*

$$t_{\widehat{\beta}_I} \Rightarrow \mathcal{Z} \tag{10}$$

Similarly, under Assumptions 1-2 and 3 (stationarity),

$$t_{\widehat{\beta}_F} \Rightarrow \frac{1}{\sqrt{1-\rho^2}}\mathcal{Z} \tag{11}$$

whereas under Assumption 3' (near-integration),

$$t_{\widehat{\beta}_F} \Rightarrow \mathcal{Z} + \frac{\rho}{\sqrt{(1-\rho^2)}} \frac{\int_0^1 \overline{J}_c(r)dW(r)}{\left\{\int_0^1 \left[\overline{J}_c(r)\right]^2 dr\right\}^{1/2}} \equiv \mathcal{A}(c,\rho) \tag{12}$$

as T is allowed to diverge.

Corollary I. *Under integration, i.e. c=0, the previous theorem holds with $\overline{W}(r)$, a demeaned Wiener process, replacing $\overline{J}_c(r)$ in* (12).

Corollary II. *Let $t_{\widehat{\beta}}$ be the t-statistic on $\widehat{\beta}$ computed from the standard (non-augmented) predictive regression (1). Then, under the same conditions as in Theorem 2.1, $t_{\widehat{\beta}_F} = t_{\widehat{\beta}}/\sqrt{1-\rho^2}+o_p(1)$, i.e., $t_{\widehat{\beta}_F}$ is asymptotically equivalent to $t_{\widehat{\beta}}$ when the former is trivially multiplied by $\sqrt{1-\rho^2}$.*

Proof. *See Appendix for details.*

Remark 2.2. It is worth commenting on the similarities and differences of the null distribution of $t_{\widehat{\beta}_F}$ with those of other statistics related to augmented predictive regressions. Amihud and Hurvich (2004) propose a heuristic finite-sample approximation of the standard error of a bias-corrected version of $\widehat{\beta}_F$ and, hence, the resulting t-statistic has a limit distribution which is different from that of $t_{\widehat{\beta}_F}$. Campbell and Yogo (2006) consider the so-called infeasible Q-statistic, defined as the t-test resulting from regressing $y_t - \beta_{uv}(x_t - \phi x_{t-1})$ on a constant and x_{t-1}. This test has optimality properties when ϕ is known and, in fact, is asymptotically equivalent to $t_{\widehat{\beta}_I}$ under this condition. When ϕ and the remaining nuisance parameters are replaced by their (consistent) sample counterparts, the resultant t-statistic, say Q_F, verifies $Q_F = t_{\widehat{\beta}_F} + o_p\left(T^{-1/2}\right)$, and so is asymptotically equivalent to $t_{\widehat{\beta}_F}$. Finally, Lewellen (2004) attempts to approach the term $\widehat{\beta} - \beta_{uv}\left(\widehat{\phi} - \phi\right) \equiv \widehat{\beta}_I$ by arbitrarily setting $\phi \approx 1$, *i.e.*, considering the worst-case scenario under stationarity restrictions. We can show that the difference $t_L - t_{\widehat{\beta}_F}$, where

t_L denotes the resulting t-test in Lewellen (2004), behaves randomly, so both tests are not equivalent.

Theorem 2.1 shows that using $\widehat{v}_t$ to make the regression approach feasible has implications when carrying out inference on β, even asymptotically. Whereas $t_{\widehat{\beta}_I}$ converges to a standard normal distribution, the large-sample distribution of its feasible counterpart $t_{\widehat{\beta}_F}$ is strongly affected by the properties of x_t. In fact, and as remarked in Corollary II, the resulting t-statistic is distributed as the t-statistic from the non-augmented predictive regression re-scaled by $\sqrt{1-\widehat{\rho}^2}$. Hence, the main practical effect is essentially to shift both the null and the alternative distribution of the LS t-ratio. Campbell and Yogo (2006) show through Monte Carlo simulation that their Bonferroni-type procedure in an augmented regression exhibits improved power over the analogous Bonferroni t-test of Cavanagh *et al.* (1995) in the standard predictive regression. Our analysis makes clear that such power gains would not arise mechanically as a consequence of simply using augmented regressions, since the distribution of Q_F is equivalent to $t_{\widehat{\beta}}/\sqrt{1-\rho^2}$. Power gains for this test stem mostly from using a more powerful and efficient unit-root test statistic (the DF-GLS test rather than the ADF test) to generate confidence intervals. Similarly, it may be possible to resort to different bias-reduction techniques as suggested by Amihud and Hurvich (2004) to enhance the relative efficiency of augmented predictive regressions.

As expected, the contemporaneous correlation ρ and the local-to-unity parameter c jointly determine the location and shape of the distributions of $t_{\widehat{\beta}_F}$ and $t_{\widehat{\beta}}$ under the set of assumptions considered. For instance, if ρ is largely negative and c is in the neighborhood of zero, $t_{\widehat{\beta}_F}$ can be largely skewed to the right, which would lead to severe overrejections when the critical values from the standard normal distribution are used. Obviously, correct inference requires precise knowledge of these nuisance parameters. As discussed previously, however, this issue poses non-trivial difficulties, since c cannot be estimated consistently. In the following section, we present a robust test that is able to deal with this problem, both in the augmented and non-augmented context. The main statistics from augmented regressions have, in principle, similar statistical properties as those from non-augmented regressions. However, in small samples, the possibility to improve power when combining this approach with further small-sample refinements cannot be ruled out. Hence, we discuss our test in the augmented-regression setting, noting that the large-sample discussion for a similar test in standard non-augmented predictive regressions follows along the same lines, after trivial rescalling.

3 A robustified test in augmented predictive regressions

3.1 Testing procedure and asymptotic null distribution

As shown in Theorem 2.1, the asymptotic properties of the feasible t-test are largely determined by the stochastic properties of the predictor, as characterized by the parameter c. Given that this term cannot be consistently estimated, it is natural to consider a test statistic that seeks to robustify the procedure against the unknown values of this nuisance parameter. In this section, we propose a modified t-statistic whose critical values under H_0 are stable with respect to the values of c in $(-\infty, 0]$, given $\rho \in (-1,1)$, and the arbitrary choice of a desired significance level, $(1-\lambda)\,\%$. To this end, the statistic $t_{\widehat{\beta}_F}$ must be re-scaled by a stochastic factor that ensures that the null cumulative distribution functions under either Assumption 3 or 3' coincide asymptotically at a chosen point, that point being the asymptotic critical value associated with the nominal significance level λ.

As discussed by Vogelsang (1998), this is possible when the re-scaling factor exhibits certain statistical properties, namely, convergence to one under stationarity $(c \to -\infty)$ and, simultaneously, be asymptotically invariant with respect to c in the most extreme case in which the predictor is allowed to exhibit a unit root $(c = 0)$. Several statistics in the unit-root literature fulfill the required properties. The interest in these statistics, is naturally justified because they convey statistical information about the order of integration of x_t and, as such, have been used in predictive analysis. For instance, unit-root test statistics play a major role in the construction of Bonferroni-type confidence intervals; see for instance, Campbell and Yogo (2006). It should be noted, however, that we just use unit-root statistics to modify the basic t-ratio for predictability and not as a pre-testing device to infer the extent of persistence of the predictor.

Thus, following Vogelsang (1998) and Harvey *et al.* (2006), we consider a simple transformation of the unit-root test statistic of Park and Choi (1988) and Park (1990)[5] which may be computed from the following auxiliary regression,

$$x_t = \gamma_0 + \sum_{i=1}^{m} \gamma_i t^i + \epsilon_{t,m}, \quad t = 1, ..., T \tag{13}$$

where x_t is the predictor variable, $m \geq 1$ is the order that characterizes the deterministic polynomial kernel used in a semi-parametric fitting of x_t, while $\epsilon_{t,m}$ denotes the regression residuals.[6] Considering (13), the following likelihood-ratio type test for the null hypothesis $\mathrm{H}_0^* : \gamma_1 = ... = \gamma_m = 0$, *i.e.*,

$$J_{x,T}\,(m) = \frac{\mathcal{RSS}_r - \mathcal{RSS}_u}{\mathcal{RSS}_u} \tag{14}$$

where $\mathcal{RSS}_u$ and $\mathcal{RSS}_r$ denote the unrestricted and restricted sums of squares of OLS residuals from (13), respectively, has power to detect unit roots in x_t against a general stationary alternative. Under Assumptions 2 and 3, $J_{x,T}\,(m) \xrightarrow{p} 0$ (cf. Park and Choi, 1988), whereas, under Assumptions 2 and 3', $J_{x,T}\,(m) = O_p\,(1)$, with a well-defined distribution that depends only on c; see Theorem 3.1 below for details. It is precisely the heterogenous behavior of the predictor under stationary and (near) integrated dynamics that we shall exploit to design a robust test in the predictive regression context. The following proposition introduces the precise way to do so.

Proposition 3.1. *Let* $t_{\widehat{\beta}_F}$ *be the feasible t-statistic defined in* (9). *Given an arbitrary probability level* $\lambda\%$, $0<\lambda<1$, *the test statistic*

$$t^*_{\widehat{\beta}_F,\lambda} = t_{\widehat{\beta}_F} \exp\left[-b^*_{\lambda\rho} J_{x,T}\,(m)\right] \tag{15}$$

with $J_{x,T}\,(m)$ *defined in* (14), *and* $b^*_{\lambda\rho}$ *denoting a specific constant value, has power to detect departures from* $H_0 : \beta = 0$ *and, for a predetermined value of* $b^*_{\lambda\rho}$, *will exhibit*

---

[5] Note that other unit root tests that share these properties are available in the literature (*e.g.*, the test statistic of Breitung, 2002) and may be used as well; see Bunzel and Vogelsang (2005).

[6] According to experimental results, Vogelsang (1998) suggests setting $m = 9$; we follow this in the experimental and empirical sections.

asymptotic nominal size $\lambda\%$ independently of whether $\{x_t\}$ is stationary or (near) integrated, given the value of ρ.

Remark 3.1. The modified t-statistic arises from re-scaling $t_{\widehat{\beta}_F}$ with the stochastic factor $\exp\left[-b^*_{\lambda\rho} J_{x,T}(m)\right]$, which depends on the unit-root statistic $J_{x,T}(m)$, and the positive, finite equalizing constant $b^*_{\lambda\rho}$. Under Assumptions 2 and 3, $J_{x,T}(m) \xrightarrow{p} 0$ and therefore $\exp\left[-b^*_{\lambda\rho} J_{x,T}(m)\right] \xrightarrow{p} 1$ for any $b^*_{\lambda\rho}$, so the scaling factor does not play any role and $t^*_{\widehat{\beta}_F,\lambda}$ is asymptotically equivalent to $t_{\widehat{\beta}_F}$. On the other hand, under Assumption 3', $\exp\left[-b^*_{\lambda\rho} J_{x,T}(m)\right] = O_p(1)$, and so the re-scaling factor introduces an additional source of randomness that changes the shape and location of the asymptotic distribution of the statistic.

Remark 3.2. The specific purpose of $J_{x,T}(m)$ is to endogenously smooth discontinuities in the limit distribution as the autoregressive root of x_t is allowed to go from the stationary to the integrated region. Since the asymptotic distribution of $t^*_{\widehat{\beta}_F,\lambda}$ in (15) is different under Assumptions 3 and 3', the purpose of the constant $b^*_{\lambda\rho}$ is to offset this divergence at a certain percentile $\lambda\%$ which, by design, will render the resultant testing procedure asymptotically size correct independently of the stochastic properties of the predictor. In other words, although the null asymptotic distribution of $t^*_{\widehat{\beta}_F,\lambda}$ is not generally invariant with respect to $(c,\rho)'$, there exists a constant value $b^*_{\lambda\rho}$ for any arbitrary nominal level $\lambda\%$ and value of ρ which is able to bring the stationary and near-integrated distributions of $t^*_{\widehat{\beta}_F,\lambda}$ close together at the λ-quantile. Although the resulting critical value still depends on ρ, this is not particularly problematic as we can identify this term consistently from the data. Note that $b^*_{\lambda\rho}$ will be different for any target probability λ and value of ρ and, in general, will be data-dependent.

The asymptotic distributions of $t^*_{\widehat{\beta}_F,\lambda}$ under Assumptions 1-2, and either Assumption 3 or 3', are characterized in the following theorem.

Theorem 3.1. *Under Assumptions 1-3 and the null hypothesis $H_0 : \beta = 0$, it follows as $T \to \infty$ that*

$$t^*_{\widehat{\beta}_F,\lambda} \equiv t_{\widehat{\beta}_F} \Rightarrow \mathcal{N}\left(0, \frac{1}{1-\rho^2}\right) \tag{16}$$

so $t^*_{\widehat{\beta}_F,\lambda}$ is asymptotically equivalent to $t_{\widehat{\beta}_F}$, whereas under Assumption 1, 2 and 3',

$$t^*_{\widehat{\beta}_F,\lambda} \Rightarrow \mathcal{A}(c,\rho)\exp(-b^*_{\lambda\rho}\mathcal{J}_c(m)) \equiv \mathcal{A}_{\lambda,m}(c,\rho) \tag{17}$$

with $\mathcal{A}(c,\rho)$ as defined in Theorem 2.1, and

$$\mathcal{J}_c(m) = \left(\int_0^1 \left[\mathcal{V}^U_{c,m}(r)\right]^2 dr\right)^{-1} \times \left(\int_0^1 \left[\mathcal{V}^U_{c,m}(r)\right]^2 dr - \int_0^1 \left[\mathcal{V}^R_c(r)\right]^2 dr\right) \tag{18}$$

with $\mathcal{V}^U_{c,m}$ and $\mathcal{V}^R_c$ denoting the residuals from the projection of $x(r)$ onto the space spanned by $[1, r, r^2, ..., r^m]$, and a constant, respectively.

Proof. *See Appendix for details.*

Theorem 3.1 provides the basis to design quantile invariant inference. For instance, assume that one is interested in testing $\mathrm{H}_0 : \beta = 0$ against a right-tailed alternative

$\mathrm{H}_1 : \beta > 0$ at a $\lambda\%$ nominal size. According to Theorem 3.1, $b^*_{\lambda\rho}$ can be determined as the solution of the implicit equation

$$\Pr\left(\mathcal{Z} > \frac{z_{1-\lambda}}{\sqrt{1-\rho^2}}\right) = \Pr\left(t^*_{\widehat{\beta}_F,\lambda} \geq \xi_{1-\lambda}\right) = \lambda \tag{19}$$

where $\xi_{1-\lambda}$ denotes the corresponding asymptotic upper-tail critical value of $\mathcal{A}_{\lambda,m}(0,\rho)$, and $z_{1-\lambda}$ is the $(1-\lambda)th$ percentile of the standard normal distribution (*e.g.*, $z_{0.95} = 1.645$ for $\lambda = 0.05$). The distribution of $t^*_{\widehat{\beta}_F,\lambda}$ when x_t has an exact unit root, *i.e.*, $\mathcal{A}_{\lambda,m}(0,\rho)$, depends only on the values of ρ, but not on c, and so $b^*_{\lambda\rho}$ is completely characterized in terms of ρ. This property can be shown as a corollary to the previous theorem and is used here to identify the constant that equals the cumulative distribution functions under Assumptions 3 and 3'. Equation (19) can then be solved using numerical methods after consistently estimating the correlation parameter ρ through its sample counterpart. Given $b^*_{\lambda\rho}$, the resultant statistic is ensured by construction to exhibit approximately $\lambda\%$ asymptotic size when the innovations have correlation ρ, given any value of the local-to-unity parameter in the support $(-\infty, 0]$. Section 4.4 provides tables with values of $b^*_{\lambda\rho}$ and corresponding critical values as a function of ρ and λ for different sample lengths.

To gain insight and intuition on this issue, Figure 3.1 depicts the cumulative distribution functions (c.d.f.) of $t^*_{\widehat{\beta}_F,\lambda}$ for $\rho = -0.95$ computed based on $1,000$ Monte Carlo simulations of a sample of $T = 1,000$ observations for $\phi = 0$ and $\phi = 1$ with i.i.d normal innovations. The equalizing constant $b^*_{\lambda\rho}$ has been computed for a right-tailed test and $\lambda = 0.05$. Note that the values given to ϕ correspond to the 'extreme' cases representing a stationary white noise process ($c = -1000$) and a random walk ($c = 0$). Given that the asymptotic distribution of $t^*_{\widehat{\beta}_F,\lambda}$ is well-defined and continuous in c, as shown in Theorem 3.1, the different c.d.f. of $t^*_{\widehat{\beta}_F,\lambda}$ that would arise for values of c within these extremes will lie in the area between the two functions depicted. We can observe that the distribution of $t^*_{\widehat{\beta}_F,\lambda}$, as a function of c, is not generally invariant. The critical values depend on the particular value of this nuisance term. Nevertheless, the 95%-quantile of any of these distributions (square box in Figure 3.1) is, by construction, exactly the same, so we may rely on the same critical values and ensure valid inference without knowledge of the true value of this parameter.

[**Insert Figure 3.1 around here**]

Remark 3.3: Setting $c = 0$ to identify the equalizing constant may be seen as evocative of the strategy employed by Lewellen (2004), who sets arbitrarily $\phi \simeq 1$ to approximate numerically the unobservable bias $\widehat{\phi} - \phi$ in his analysis. Under stationarity, this leads to the most conservative assumption for testing predictability which, in the case of Lewellen's test, has the unpleasant effect of ensuring approximately correct size only if $\phi \simeq 1$; see also Campbell and Yogo (2006). In sharp contrast, imposing $c = 0$ in the testing approach to determine the equalizing constant in our procedure does not imply noticeable size distortions in the subsequent t-statistics even if the autoregressive coefficient ϕ largely enters into the stationary region, since the procedure holds for the whole support of c. This property is discussed in greater detail below in the Monte Carlo analysis of section 4.1.

Remark 3.4: Paralleling the previous discussion, we can design a robust test in the context of non-augmented regression, say $t^*_{\widehat{\beta},\lambda}$, which will exhibit similar properties. The only difference is that the equalizing constants for this test would be different. The procedure would control size exactly in the same terms and exhibit similar power performance, as discussed in the following subsection.

3.2 Asymptotic power function

In this subsection, we discuss the properties of the robust test when the true value of β departs from the null hypothesis. Considering, under stationarity, the local alternative $\beta = \beta_a/\sqrt{T}$, with some constant $|\beta_a| > 0$, we can show as a corollary of Theorem 3.1 that,

$$t^*_{\widehat{\beta}_F,\lambda} \Rightarrow \beta_a \frac{\sigma_v}{\sigma_\varepsilon\sqrt{1-\phi^2}} + \frac{1}{\sqrt{1-\rho^2}}\mathcal{Z} \tag{20}$$

whereas, under near-integration, and for alternatives of the form $\beta = \beta_a/T$, it follows that

$$t^*_{\widehat{\beta}_F,\lambda} \Rightarrow \beta_a \frac{\sigma_v \exp(-b^*_{\lambda,\rho}\mathcal{J}_c\,(m))}{\sigma_\varepsilon \left(\int_0^1 \left[\overline{J}_c(r)\right]^2 dr\right)^{-1/2}} + \mathcal{A}_{\lambda,m}\,(c,\rho) \tag{21}$$

with $\mathcal{A}_{\lambda,m}\,(c,\rho)$ and $\mathcal{J}_c\,(m)$ as defined in (17) and (18), respectively. The limit results in (20) and (21) allow us to characterize the asymptotic power function of $t^*_{\widehat{\beta}_F,\lambda}$, say $\mathcal{P}^*\,(\beta_a)$. In particular, for a right-tailed alternative and under Assumption 3, we will have

$$\mathcal{P}^*\,(\beta_a) = E\left[\bar{\Phi}\left\{z_{1-\lambda} - \beta_a \frac{\sigma_v}{\sigma_u\sqrt{1-\phi^2}}\right\}\right] \tag{22}$$

whereas under Assumption A3',

$$\mathcal{P}^*\,(\beta_a) = E\left[\bar{\Phi}\left\{\frac{z_{1-\lambda}\exp(b^*_{\lambda\rho}\mathcal{J}_c\,(m))}{\sqrt{(1-\rho^2)}} - \mathcal{B}\,(c,\rho) - \beta_a\frac{\sigma_v}{\sigma_\varepsilon}\left(\int_0^1 \left[\overline{J}_c(r)\right]^2 dr\right)^{1/2}\right\}\right] \tag{23}$$

where the expectation is taken over the distribution of $(J_c(r), W\,(r))'$, given $\bar{\Phi}\,\{\tau\} = 1 - \Pr\,(\mathcal{Z} \leq \tau)$, and the bias term $\mathcal{B}\,(c,\rho) = \mathcal{A}\,(c,\rho) - \mathcal{Z}$. Consequently, the test exhibits non-trivial asymptotic power against local alternatives, *i.e.*, for fixed values $\beta = \beta_a$, $\mathcal{P}^*\,(\beta_a) \to 1$ as T diverges, so the procedure is consistent. Note that, since the rate of convergence of $\widehat{\beta}_F$ under near-integration is T rather than $\sqrt{T}$, the robust test is expected to be more powerful in the near-integrated region, a property also shared by other tests in this context.

Let $\mathcal{P}^*_n\,(\beta_a)$ denote the power functions of the robust t-test, $t^*_{\widehat{\beta},\lambda}$, computed in a non-augmented predictive regression. Under stationarity, it is immediate to show that $\mathcal{P}^*\,(\beta_a) = \mathcal{P}^*_n\,(\beta_a) = \mathcal{P}_s\,(\beta_a)$ for all β_a, where $\mathcal{P}_s\,(\beta_a)$ is the power function of the standard LS t-test in a non-augmented regression, so the power function is the same for the three statistics. Similarly, in the near-integrated context, we can show that

$$\mathcal{P}^*_n\,(\beta_a) = E\left[\bar{\Phi}\left\{\frac{z_{1-\lambda}\exp(b^{*na}_{\lambda\rho}\mathcal{J}_c\,(m))}{\sqrt{(1-\rho^2)}} - \mathcal{B}\,(c,\rho) - \frac{\beta_a\sigma_v}{\sigma_\varepsilon}\left(\int_0^1 \left[\overline{J}_c(r)\right]^2 dr\right)^{1/2}\right\}\right] \tag{24}$$

where $b^{*na}_{\lambda\rho}$ denotes the non-augmented counterpart of $b^*_{\lambda\rho}$, *i.e.*, the value of the equalizing constant for the modified t-test from the non-augmented regression, namely, the implicit solution to the equation

$$\Pr\left(t^*_{\widehat{\beta},\lambda} \geq z_{1-\lambda}\right) = \lambda. \tag{25}$$

Clearly, the power functions of both tests are the same if $b^*_{\lambda\rho} = b^{na}_{\lambda\rho}$. Because $t^*_{\widehat{\beta}_F,\lambda} = \sqrt{(1-\rho^2)}t^*_{\widehat{\beta},\lambda}$, then $\Pr\left(t^*_{\widehat{\beta}_F,\lambda} \geq x\right) = \Pr\left(t^*_{\widehat{\beta},\lambda} \geq \sqrt{(1-\rho^2)}x\right)$, and hence $b^*_{\lambda\rho} = b^{na}_{\lambda\rho}$. Consequently, the distributions of the robust tests in the augmented and non-augmented regressions only differ in the scaling factor $\sqrt{(1-\rho^2)}$, both under the null and the alternative hypotheses and, therefore, both tests are expected to exhibit similar asymptotic performance.[7]

Nevertheless, and as noted by Campbell and Yogo (2006), it may be possible to obtain relative efficiency gains by using suitable estimation refinements which are enabled in the augmented regression framework. A first possibility is to use bias-reduction techniques on the autoregressive coefficient, as proposed by Amihud and Hurvich (2004). The bias reduction may then feed into the estimation of β and ensure more precise estimates, thereby increasing power. Secondly, it is possible to use estimation techniques of the deterministic component in the predictor which avoid inflating the autoregressive bias, such as for instance recursive demeaning (see Shin and So, 2001). Simulations (available upon request) showed that these deliver some improvements. In the Monte Carlo section, we will analyse the basic performance of our test without the application of any bias reduction approaches.

3.3 Generalizing the basic assumptions

Assumption 2 seems particularly restrictive for practical purposes given the stylized features that characterize financial and economic data. We can instead consider the following generalization that allows the innovations in equations (1)-(2) to exhibit short-run dynamics and/or (conditional) heterogeneity, and which does not impose any particular distributional restriction on the data. In the sequel, given a generic stochastic process η_t, we use $\mathcal{F}^\eta_t$ to denote its filtration, *i.e.*, $\mathcal{F}^\eta_t = \sigma\{\eta_s, s \leq t\}$.

Assumption 2'. *Let* $A(L) = 1 - \sum_{j=1}^{p} \delta_j L^j$, $p \geq 0$, *be a p-th order autoregressive filter in the lag-operator L with fixed roots and* $A(z) \neq 0$ *for all* $|z| \leq 1$. *Then:*
i) $\{u_t, \mathcal{F}^u_t\}$ *is a strictly stationary and ergodic Martingale Difference Sequence (MDS) with* $E(u^4_t) < \infty$.
ii) $v_t = A^{-1}(L)\, e_t$, *where* $\{e_t, \mathcal{F}^e_t\}$ *is a strictly stationary and ergodic MDS with unconditional variance* σ^2_e, *and* $E(e^4_t) < \infty$.
iii) Let $\xi_t = (u_t, e_t)'$. *Then,* $E[\xi_t\xi'_t] = \Omega_{ue} \equiv [\sigma^2_u, \sigma_{ue}; \sigma_{ue}, \sigma^2_e]$, *where* $E(e^2_t) \equiv \sigma^2_e > 0$, $E(u^2_t) \equiv \sigma^2_u > 0$, $\sigma_{ue} = \rho\sigma_u\sigma_e$ *for some* $\rho \in (-1, 1)$.

Under these general conditions, the predictor may be serially correlated and have up to one unit root, with dynamics given by $A(L)(1 - \phi L)(x_t - \mu^*) = e_t$. The short-run component may be driven by a stationary AR(p) process with innovations obeying a MDS, which in turn enables the conditional variance of the process to follow, among others, (stationary) GARCH-type and stochastic volatility patterns. Other general restrictions, such as mixing conditions, could alternatively be considered without significantly changing any of the results, provided that the partial sums of ξ_t satisfy a functional central limit theorem. It should be noted that Assumptions 3 and 3' still determine the long-run dynamics of the predictor and, hence, the underlying properties of the main statistics. Under Assumptions 1, 2' and 3, x_t is driven by a stationary AR$(p+1)$ process. Under conditions 1, 2' and 3', the largest root in the autoregressive representation of x_t is near unity, whereas the short-run dynamics is driven by a stationary AR(p) model. Note that, for practical purposes, the short-run dynamics may be characterized by a stationary and invertible linear process such that the AR(p) model, for some large enough $p < \infty$, approaches the underlying AR representation reasonably well.

---

[7] This has also been confirmed through Monte Carlo simulations. Results can be obtained from authors upon request.

The presence of contemporaneous correlation $\rho \neq 0$ of the innovations of the two-equation system keeps playing a crucial role because of the simultaneity problem when the largest autoregressive root is close to unity. It is necessary to identify this term in order to devise correct inference, noting that ρ can be consistently estimated through the sample correlation of the OLS residuals $\widehat{e}_t = x_t - \widehat{\tau} - \sum_{i=1,p+1} \widehat{\pi}_i x_{t-i}$ and $\widehat{u}_t = y_t - \widehat{\alpha} + \widehat{\beta} x_{t-1}$. Because $u_t = \beta_{ue} e_t + \varepsilon_t$, with $\beta_{ue} \equiv \sigma_{ue}/\sigma_e^2$. Under Assumptions 1 and 2', where ε_t now behaves as a MDS, we could still attempt to reduce part of the variability of y_t which is uncorrelated with x_{t-1} through the (feasible) augmented regression

$$\left[y_t - \widehat{\beta}_{ue} \widehat{e}_t \right] = \alpha + \beta x_{t-1} + \varepsilon_t \tag{26}$$

to ensure more powerful testing, where $\widehat{\beta}_{ue}$ is the slope estimate of the regression of $\widehat{u}_t$ on $\widehat{e}_t$. The following theorem discusses formally the limit distributions of the resultant t-stastistic under the set of general conditions considered.

Theorem 3.2. *Let $t_{\hat{\beta}_F,p}$ be the resulting t-statistic of the estimated slope coefficient in the augmented regression* (26) *. Under the null hypothesis and Assumptions 1 and 2', it follows, as $T \to \infty$, that under Assumption 3*

$$t_{\hat{\beta}_F,p} \Rightarrow \frac{1}{\sqrt{1-\rho^2}} \mathcal{Z} \tag{27}$$

and under Assumption 3',

$$t_{\hat{\beta}_F,p} \Rightarrow \mathcal{A}\left(c, \rho\right) \tag{28}$$

with $\mathcal{A}\left(c, \rho\right)$ defined as in Theorem 2.1.

Proof. *See Appendix for details.*

Remark 3.5. The main implication of Theorem 3.2 is that $t_{\hat{\beta}_F,p}$ is asymptotically distributed as $t_{\hat{\beta}_F}$. Hence, the main results discussed in Section 2 hold directly if the innovations in the predictor are uncorrelated but not necessarily independent. Similarly, joint normality is not required. The existence of short-run dependence in the innovations of the predictive variable is essentially handled in the same spirit as the long-run AR(1) dynamics, namely, using augmentation to obtain the empirical proxy of the MDS innovations. This technique is similar to the approach used in the well-known ADF unit-root test. As in that test, the resulting test statistic is scale-invariant and, therefore, overrides the need of estimating long-run variance parameters. In practice, the lag-order of the autoregressive process is not required to be known and can be determined based on data-dependent procedures (such as AIC or BIC, for instance). Because of the asymptotic equivalence, we can use the same approach introduced previously to ensure valid inference in this context. The properties of such statistics are discussed in the following Theorem.

Theorem 3.3. *Let $t_{\hat{\beta}_F,p}$ be the t-statistic in Theorem 3.2, and define*

$$t^*_{\widehat{\beta}_F,p} = t_{\hat{\beta}_F,p} \exp\left[-b^*_{\lambda\rho} J_{x,T}(m)\right]$$

a modified t-statistic in the spirit of Theorem 3.1, with $J_{x,T}(m)$ defined in (14), and the same equalizing constant $b^*_{\lambda,\rho}$ as in Theorem 3.1. Then, under the same conditions as in Theorem 3.2, $t^*_{\widehat{\beta}_F,p}$ is asymptotically distributed as $t^*_{\widehat{\beta}_F,\lambda}$ in Theorem 3.1.

Proof. *See Appendix for details.*

These results show that the values of the equalizing constant $b^*_{\lambda\rho}$ and the corresponding critical values related to the distribution discussed in the i.i.d. context and which are applicable to $t^*_{\widehat{\beta}_F,\lambda}$ are still valid for the 'whitened' $t^*_{\widehat{\beta}_F,p}(\lambda)$ statistic under the more general conditions studied here. Given that these results are obtained under large-sample theory, the following section analyzes the small-sample behavior of the main statistics involved.

4 Experimental analysis

Throughout the following subsections, we analyze by means of Monte Carlo simulations different issues related to the finite-sample behavior of the robustified t-test $t^*_{\widehat{\beta}_F,\lambda}$ described in the previous section. Firstly, we investigate whether inference based on the robustified test ensures approximately correct empirical size in small samples with independence of the values of the nuisance parameter c. Secondly, we evaluate the power performance of the procedure, and compare it to alternative procedures, when the predictive variable is driven by either i.i.d. innovations or exhibits short-run dynamics. Finally, the last experimental section provides the necessary equalizing constants and critical values of the distribution as a function of the estimated ρ parameter for different sample lengths and significance levels.

4.1 Size analysis in small samples

As discussed in Section 3, implementing valid inference in the context of predictive regressions implies dealing with the unknown parameters $(c, \rho)'$ that characterize the asymptotic distribution. The basic approach to conduct inference on the basis of the robustified test is to first infer ρ (through its OLS counterpart) and, given this estimate and the choice of the significance level, to determine the suitable values of the equalizing constant $b^*_{\lambda\rho}$ and the corresponding critical value; see also Vogelsang (1998). In this context, therefore, the most important question refers to whether $t^*_{\widehat{\beta}_F,\lambda}$ exhibits approximately correct size, particularly, when the available sample is small, as asymptotic theory is expected to provide good large sample approximations.

To address this question, we consider a one-sided test of predictability for $\mathrm{H}_0 : \beta = 0$ vs $\mathrm{H}_1 : \beta > 0$ at a nominal level of $\lambda = 5\%$. We chose this framework because it is common in the empirical literature to consider right-sided alternatives, since finance theory often suggests $\beta \geq 0$ for many predictors, such as, for instance, the dividend yield. The main aim of this experiment is to evaluate the average frequency of rejection of the test when the true predictive coefficient is $\beta = 0$.

The design of the Monte Carlo experiment is as follows. We consider $5,000$ replications of a sample of length $T = 100$ of the two-equation data generating process (1)-(2) setting $\beta = 0$. The set of innovations $\xi_t = (u_t, v_t)'$ in this system is driven from an i.i.d. Gaussian bivariate distribution with zero mean and covariance matrix having normalized diagonal entries $\sigma_u^2 = \sigma_v^2 = 1$ and correlation parameter $\rho \in \{0, -0.8 - 0.9, -0.95\}$. This set includes empirically relevant values that range from uncorrelated innovations, as in the case of interest rates, to strongly contemporaneous negative correlations, as is typically observed when considering valuation ratios. The intercepts α and μ in equations (1) and (2) are set to zero, and so are all initial conditions used to initialize the autoregressive process. Finally, we focus on the general setting provided by Assumption 3' and assume an autoregressive coefficient of the form $\phi = 1 + c/T$, with c taking a continuous sequence of values in the interval $[-50, 0]$. This set contains values that imply autoregressive dynamics ranging from stationarity ($\phi = 0.5$) to integration ($\phi = 1$). The robust test statistic $t^*_{\widehat{\beta}_F,\lambda}$ is computed for each of these simulations using augmented predictive regression (6), with the values of the equalizing constant chosen according to the LS-based estimate $\widehat{\rho} = \sum_{t=2}^{T} \widehat{u}_t \widehat{v}_t / T$. For comparative purposes, we also consider the naive LS t-statistic for $\widehat{\beta}$ computed from the non-augmented predictive regression (1), with critical values corresponding to the standard normal distribution for all c and ρ. This test is expected to largely overreject as c approaches zero and $|\rho|$ increases, whereas the test proposed in this paper is expected to show a frequency of rejection around the 5% level in all cases. The average frequencies of rejection of these two tests are displayed in Figure 4.1.

[**Insert Figure 4.1 around here**]

The four panels in Figure 4.1 clearly show that, whereas the standard t-statistic suffers considerable size distortions when ρ approaches its lower bound and c tends to zero, the modified robust t-statistic shows approximately correct size around the 5% level in all cases analyzed. According to our simulations, the empirical size of this test lies in the range $[4.75\%, 6.05\%]$, which seems a fair price for the flexibility and tractability ensured in this context. As noted previously, other alternative procedures may have problems to control size properly uniformly on the support of c. For instance, the test in Lewellen (2004) has optimal asymptotic properties if the true value of c is close to (but strictly less than) one, but can be largely undersized otherwise; see, for instance, the experimental analysis reported in Campbell and Yogo (2006). Similarly, the experimental simulations in Jansson and Moreira (2006) report under the same Monte Carlo setting considered here that the Bonferroni test of Campbell and Yogo (2006) is distorted when c departs largely from the origin. The reason is that the test is designed to have asymptotic size equal to 5% when c is bounded in the range $[5, -50]$. For instance, these authors report an empirical size of 0.7% for $c = -50$ and $\rho = -0.5$ when $T = 100$, and even larger distortions for other parametric configurations of the vector $(c, \rho)'$. In sharp contrast, our testing procedure has empirical size approximately equal to the expected asymptotic level in all these cases.

4.2 Power analysis: independent and weakly correlated innovations

In this section, we evaluate the empirical power of the robustified test given a data generating process that allows the predictor to be driven by either i.i.d. or weakly-dependent innovations. As in the previous experiments, we test predictability against a right-tailed alternative. For comparative purposes, we analyze the performance of the modified t-test in relation to other valid testing procedures in this setting. It should be noted that, in the general context covered by the near-integration theory, there is no uniformly most powerful test (Elliott, Rothenberg and Stock, 1996), and the performance of different procedures is completely data-dependent. Our main purpose, therefore, is to offer a comparative reference to appraise the relative power of our test and validate its empirical suitability rather than to conduct a horse race.

As benchmarking methods, we consider the Bonferroni t-test of Cavanagh *et al.* (1995), which builds on the standard (non-augmented) predictive regression, the Bonferroni Q-test of Campbell and Yogo (2006), which builds on an augmented predictive regression, and the robust test of Maynard and Shimotsu (2009), which is an alternative non-parametric approach to regression-based tests. Our interest in the latter is further motivated by the fact that this test will exhibit correct asymptotic size independently of all $c \in (-\infty, 0]$. In the sequel, we briefly sketch the main features of these tests.[8]

Cavanagh *et al.* (1995) proposed a Bonferroni method to make the t-test in the predictive regression feasible, which essentially consists of constructing a $100\,(1-\lambda_1)\,\%$ confidence interval for ϕ, say $\mathcal{C}_{\phi}\left(\lambda_1\right)$, and then a $100\,(1-\lambda_2)\,\%$ confidence interval for β for any $\phi \in C_{\phi}\left(\lambda_1\right)$, say $\mathcal{C}_{\beta|\phi}\left(\lambda_2\right)$. By Bonferroni's inequality, the region $\cup\mathcal{C}_{\beta|\phi}\left(\lambda_2\right)$ is a $100\,(1-\lambda_1-\lambda_2)\,\%$ confidence level for β that does not depend on ρ. Following these authors, we construct a one-sided 95% confidence interval for β, using the Dickey-Fuller unit-root test to construct the confidence interval for ϕ, and the asymptotic valid size-adjusted refinement method described therein. Similarly, the Bonferroni Q-test proposed in Campbell and Yogo (2006) attempts to make the t-ratio $t_{\widehat{\beta}_F}$ in augmented predictive regressions feasible using a completely similar strategy, but further improvements are introduced in this test. Following these authors, a one-sided 95% confidence interval is constructed, using the DF-GLS unit-root test to define a confidence interval for ϕ, and then using the refined procedure described therein to make the resultant test less conservative.

Finally, the non-parametric test in Maynard and Shimotsu (2009) ensures robust inference independently of c. Note that, whereas our test can achieve this property for an arbitrary quantile of the cumulative distribution function, the nonparametric test ensures invariance over the whole distribution. The main statistic is denoted here as Λ_T, with $\Lambda_T = \sum_{h=1}^{T-1} k\left(\tau\right)\Gamma_{\tilde{y}\Delta x}\left(h\right)$, $\tau = \left(h-1\right)/m_T$, $\Gamma_{\tilde{y}\Delta x}\left(h\right) = T^{-1}\sum_{t=h+1}^{T}\tilde{y}_t\Delta x_{t-h}$, where m_T is a bandwidth and $k(\tau)$ a kernel function satisfying the usual conditions. Under the null hypothesis of no predictability, $\sqrt{T/m_T}\Lambda_T \Rightarrow \mathcal{N}\left(0,\varsigma\right)$ as T is allowed to diverge, where $\varsigma > 0$ is the limit variance. When the innovations u_t are assumed to follow a MDS, ς can be consistently estimated as

$$\frac{1}{m_T}\sum_{h'=1}^{T-1}\sum_{h=1}^{T-1} k\left(\tau\right)k\left(\tau'\right)k_2\left(\frac{h'-1}{m_{2T}}\right)\Gamma_{\Delta x\Delta x}\left(h'-h\right)\Gamma_{\tilde{y}\tilde{y}}\left(h'-h\right) \tag{29}$$

where $\tau' = \left(h'-1\right)/m_T$, $m_{2T} = o\left(m_T\right)$ is an additional bandwidth parameter, and $k_2\left(\tau\right)$ is a kernel. To implement this test, we follow Maynard and Shimotsu (2009) and consider the Bartlett kernel for both $k(\tau)$ and $k_2(\tau)$, determine m_T automatically through data-dependent techniques and set $m_{2T} = \left[(m_T)^{0.9}\right]$.

---

[8] We have also evaluated the performance of other tests such as the KPSS test (Lanne, 2002) and the infeasible t-test (Valkanov, 2003), but given that the behaviour of these tests was always inferior when compared to the behaviour of the procedures analysed, we opted for omitting these results (these can however be obtained from the authors upon request). For the application of the infeasible t-test we assumed that c is known and computed the respective critical values under this assumption.

We discuss the relative performance of all these tests to reject a sequence of local alternatives of the form $\beta = T^{-1}\beta_a$, with β_a in the interval $[0, 20]$. The design of the Monte Carlo simulations is the same as in Sections 4.1 and 4.2, but $\{x_t\}$ is now allowed to be driven by a more general class of DGPs, namely, $(1 - \phi L)(1 - \delta L)x_t = v_t$, with either $\delta = 0$ corresponding to the case of i.i.d. increments (Assumption 2), or $\delta = 0.5$ corresponding to first-order stationary dynamics (Assumption 2'). We characterize the asymptotic power function of the tests against the local alternative with $T = 500$. Figure 4.2 displays the average frequencies of rejection for different values of the nuisance vector $(c, \rho)'$ and $\delta = 0$.

[**Insert Figure 4.2**]

The main conclusion from this analysis is that the robust t-test not only presents correct size properties but also shows good size performance for empirically relevant values. For $c = -5$, which corresponds to $\phi = 0.99$, the robust t-test and the refined Bonferroni Q test show similar performance, with both tests largely dominating the Bonferroni t-test and the non-parametric test. Because the rate of convergence of the non-parametric test depends on the bandwidth, it has a much slower rate of convergence than the parametric tests, and consequently is considerably less powerful in all cases analyzed. However, it should be noted that it may show robustness properties under model misspecification or more general conditions than the other alternatives explored here.

For a smaller degree of persistence as measured by c, and/or endogenous correlation as measured by $|\rho|$, the robust t-test dominates both the Bonferroni tests and the non-parametric test in our simulations. However, it should be noted that the Bonferroni Q test seems to dominate the other alternatives when c is in the neighborhood of zero, which suggests complementary properties. When the degree of endogenous correlation $|\rho|$ decreases, the procedures in the augmented regression lose efficiency in relation to the standard regression, as expected from the theoretical discussion in Section 3.2. This is confirmed for $c = -5$ and $\rho = -0.75$. The Bonferroni Q-test and, particularly, the robust t-test, still do a good job and largely dominate the other tests. When $\rho = 0$ (not reported here), all these tests show the same performance. Also, when the degree of persistence decreases, all parametric tests lose efficiency to reject the false null in relation to the procedures in the standard regression setting, as expected from the different rate of convergence. When $c = -20$, which corresponds to a slow mean-reverting process with AR coefficient of $\phi = 0.965$, the power function of the Bonferroni Q-test tends to collapse around that of the Bonferroni t-test, and the robust t-test strongly dominates the other alternatives. In summary, the overall evidence suggests that the robust t-test seems to be a good procedure.

[**Insert Figure 4.3**]

Finally, Figure 4.3 presents the results of the Monte Carlo experiments when the DGP includes short-run dynamics with an autoregressive parameter $\delta = 0.5$. Because the order of short-run dynamics in the autoregressive model is unknown in practical settings, we adopt the following strategy to obtain a picture of the power under more realistic conditions. We consider a maximum lag-order $\bar{p} = 6$ and then use the Bayesian Information Criteria to identify the most parsimonious model to be used in the case of our test. The Bonferroni procedures use non-parametric estimates of the long-run variance of the AR process, while the non-parametric tests still use the same approach. It should firstly be remarked that the empirical size of the robust t-test $t^*_{\widehat{\beta}_F,p}(\lambda)$ ranges from 5.7% to 6.39%, showing a mild oversizing effect from considering additional lags in the first-stage regression. Therefore, the presence of short-run dynamics can successfully be accommodated through augmentation and, as discussed in Section 3.4, the distribution discussed under i.i.d. still provides a good representation. As shown in Figure 4.3, the main conclusions are similar to those obtained for the i.i.d. case. The most relevant effect is that the existence of short-run dynamics in the predictor improve the power performance of all tests analyzed. Consequently, the robust test tends to exhibit good power performance, both in absolute terms, and in relation to the alternative procedures.

4.3 Values of $b^*_{\lambda,\rho}$ and critical values as a function of ρ

One of the main advantages of the procedure introduced in (15) is that the critical values for the modified t-statistic at a nominal size of λ% may readily be computed by straightforward simulation under the null hypothesis of no predictability, $\mathrm{H}_0 : \beta = 0$. Table 4.1 presents several finite-sample and asymptotic upper tail null critical values for this test as a function of the degree of contemporaneous correlation ρ. Table 4.2 presents the (asymptotic) values of the corresponding equalizing constants. In the simulations, the dependent variable was generated under the null hypothesis, $y_t = u_t$, and given that the null critical values do not depend on c nor on other population characteristics, we may directly set $x_t = v_t$. The innovations $(u_t, v_t)'$ were simulated according to a multivariate Normal distribution with zero mean vector and covariance matrix with unit diagonal entries and off-diagonal elements characterized by ρ. Following standard practices, we considered $T = 1,000$ to approach the asymptotic distribution.

[**Insert Table 4.1 and 4.2 about here**]

Note, from Table 4.1, that the standard normal distribution may be a reasonable distribution in the context of relatively mild negative contemporaneous correlation, *i.e.*, for $\rho \leq -0.5$. However, and as expected from the theoretical results, the relevant critical values largely depart from those of the standard normal distribution for larger levels of the correlation parameter.

5 Empirical Application

To illustrate empirically the performance of the tests introduced in this paper, we analyze the S&P 500 index data as well as the annual, quarterly and monthly NYSE/AMEX value-weighted index data from the CRSP database analyzed in Campbell and Yogo (2006).[9] Our objective here is to illustrate the usefulness of our testing method and compare the results with those obtained in the original study under an alternative testing procedure, as we have discussed that both tests may complement each other and seem to exhibit better power properties than other alternative procedures. The data covers the period from 1880-2002 and we analyze predictability in three periods: *i*) the full sample 1880-2002 for all variables considered, *ii*) the subsample from 1880 to 1994 also for all variables, and finally *iii*) a subsample from 1952 to 2002 only for the CRSP data. The main predictors are the earnings-price and the dividend price ratios. The interest in the latter period lies in that it allows for the analysis of two additional predictor variables, namely the three-month T-bill rate and the long-short yield spread. Note that Campbell and Yogo (2006) suggest a testing procedure to formally determine whether a conventional t-statistic could deliver reliable inference. Because our testing ensures correct size independently of the extent of persistence, the need of conducting pre-testing on the properties of the predictor is overridden.

---

[9] For a detailed description of the data and construction of the variables see Campbell and Yogo (2006, pp. 45-46). We thank the authors for making those data available.

Table 5.1 presents the results of the modified t-statistic, for $\widehat{\beta}$ estimated from the conventional predictive regression (1), and $\widehat{\beta}_{F,p}$ computed from the modified augmented regression, for the empirical estimate of the correlation coefficient (ρ) and the order of augmentation (p) used to capture the dynamics of the predictor variables. The results presented in this section can be directly compared with those obtained by Campbell and Yogo (2006).

[**Insert Tables 5.1 about here**]

When considering the full sample, Campbell and Yogo (2006) reject the null of no predictability for the earnings-price at all frequencies and for the dividend ratio on the basis of the Bonferroni Q-test. These authors also apply the Bonferroni t-test, noting that it never rejects the null hypothesis of no predictability. From Table 5.1, we observe that our testing procedure rejects the null hypothesis, thereby reaching the same qualitative conclusions about the existence of predictability as those reported in Campbell and Yogo (2006).

In the subsample through 1994 the results are qualitatively similar. In particular, the Bonferroni Q-test and our modified t-test find predictability with the earnings-price ratio at all frequencies. The Bonferroni t-test also finds predictability in this sub-sample. The concordance of all these procedures suggests that the evidence for predictability is particularly strong in this period. Finally, in the sub-sample starting from 1952, we cannot reject the null hypothesis for valuation ratios, but we find predictability in terms of the T-bill rate and the yield spread at all frequencies except at the annual frequency. Hence, from this limited empirical analysis a similar conclusion as in Campbell and Yogo (2006) can be drawn, confirming the evidence of a predictable time-varying expected return component.

6 Conclusion

A large number of tests for predictability have been proposed in the literature over the last two decades. In particular, attention has been given to two main characteristics: *i*) the strong negative correlation between the predictive regression errors and the errors of the process underlying the generation of the predictor (which is typically assumed to be an autoregressive process) and *ii*) the near integrated (strong persistent) regressors used in predictive regressions. In this paper we introduced a modified t-statistic for predictability computed from an augmented predictive regression in which the dependant variable is stationary and the predictor is allowed to be either stationary, near-integrated or integrated.

Our procedure ensures valid inference with approximately correct size in small samples and shows good power properties for empirical applications. Correct size is ensured by construction, whereas augmented predictive regressions provide improved power over alternative procedures. Together with its methodological simplicity, these properties make the test introduced in this paper relevant for empirical applications. Our test shows two appealing features: it builds upon a variance-reduction technique that ensures more efficient estimates (and hence, has better power properties), and ensures approximately correct size independently of the extent of persistence and the stochastic properties of the predictor. Furthermore, the theoretical setting studied in this paper can readily be extended to account for more general data generating processes, such as long-memory dynamics in the predictor variable, and parameter instability, which raise interesting questions to be analyzed in future research.

References

[1] Amihud, Y. and C.M. Hurvich (2004) Predictive Regressions: A Reduced-Bias Estimation Method, *Journal of Financial and Quantitative Analysis* 39, 813-841.

[2] Amihud, Y., C.M. Hurvich and Y. Wang (2009) Multiple-Predictor Regressions: Hypothesis Testing, *Review of Financial Studies* vol. 22(1), 413-434.

[3] Breitung, J. (2002) Nonparametric Tests for Unit Roots and Cointegration, *Journal of Econometrics* 108, 343-363.

[4] Bunzel, H. and T.J. Vogelsang (2005) Powerful Trend Function Tests That Are Robust to Strong Serial Correlation, With an Application to the Prebisch-Singer Hypothesis, *Journal of Business & Economic Statistics* 23, pp. 381-394.

[5] Campbell, J.Y. (2008) Viewpoint: Estimating the equity premium. *Canadian Journal of Economics* 41(1), 1-21.

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

[7] Campbell, J.Y. and R.J. Shiller (1988) Stock Prices, Earnings, and Expected Dividends, *Journal of Finance* vol. 43(3), 661-76.

[8] Caporale, G.M. and N. Pittis (1999) Unit Root Testing Using Covariates: Some Theory and Evidence. *Oxford Bulletin of Economics and Statistics* 61(4), 583-95.

[9] Cavanagh, C. L., G. Elliott, and J. H. Stock, (1995) Inference in Models with Nearly Integrated Regressors, *Econometric Theory* 11, 1131–1147.

[10] Davidson, J. (1994) *Stochastic Limit Theory*, Oxford University Press.

[11] Elliott, G., and J. H. Stock, (1994) Inference in Time Series Regression when the Order of Integration of a Regressor is Unknown, *Econometric Theory* 10, 672–700.

[12] Elliott, G., T.E. Rothenberg and J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813 - 836.

[13] Fama, E., French, K., (1988). Dividend yields and expected stock returns. *Journal of Financial Economics* 22, 3–25.

[14] Fuller, W.A. (1996) *Introduction to statistical time series.* Wiley.

[15] Hansen, B. (1995) Rethinking the univariate approach to unit root tests: How to use covariates to increase power. *Econometric Theory* 11, 1148-1171.

[16] Harvey, D.I., S.J. Leybourne and A.M.R. Taylor (2006). Modified tests for a change in persistence. *Journal of Econometrics* 134, 441–469.

[17] Jansson, M. and M. Moreira (2006) Optimal Inference in Regression Models with Nearly Integrated Regressors, *Econometrica* 74, 681-714.

[18] Lanne, M., (2002) Testing the predictability of stocks returns. *Review of Economics and Statistics* 84, 407–415.

[19] Lewellen, J., (2004) Predicting Returns with Financial Ratios, *Journal of Financial Economics* 74, 209-235.

[20] Maynard, A. and K. Shimotsu (2009) Covariance-based orthogonality tests for regressors with unknown persistence, *Econometric Theory* 25, 63-116.

[21] Mankiw, N. G., and M. D. Shapiro (1986) Do We Reject Too Often? *Economics Letters* 20, 139–145.

[22] Park, J. Y. (1990), Testing for Unit Roots and Cointegration by Variable Addition, in T. Fomby and F. Rhodes (eds), Advances in Econometrics: Cointegration, Spurious Regressions and Unit Roots, London: Jai Press, pp. 107-134.

[23] Park, J. Y. and Choi, I. (1988), A New Approach to Testing for a Unit Root, Center for Analytic Economics, Cornell University, Working Paper 88-23.

[24] Phillips, P.C.B. (1987) Towards a unified asymptotic theory for autoregression, *Biometrika* 74, 535-547.

[25] Phillips, P. C. B. (1988), Regression theory for near-integrated time series, *Econometrica* 56, 1021-1043.

[26] Rodrigues, P.M.M. and Rubia A. (2011) Predictive Regressions: To augment or not to augment that is the question. Mimeo.

[27] Shin, D.W. and B.S. So (2001) Recursive mean adjustment for unit root tests, *Journal of Time Series Analysis* 22, 595-612.

[28] Stambaugh, R.F. (1986) Bias in regressions with lagged stochastic regressors, CRSP Working Paper No 156, University of Chicago.

[29] Stambaugh, R. (1999) Predictive Regressions, *Journal of Financial Economics* 54, 375-421.

[30] Valkanov, R. (2003) Long-horizon regressions: theoretical results and applications, *Journal of Financial Economics* 68, 201-232.

[31] Vogelsang, T. J. (1998) Trend Function Hypothesis Testing in the Presence of Serial Correlation Parameters, *Econometrica* 65, 123-148.

[32] White, H. (2001). *Asymptotic Theory for Econometricians*, Academic Press.

Technical Appendix

Given $y = (y_1, ..., y_T)'$, denote $y^* = y - E(u|v)$, with $u = (u_1, ..., u_T)'$ and $v = (v_1, ..., v_T)'$, noting that $\varepsilon = u - \beta_{uv} v$. Let X be the $(T \times 2)$ matrix of regressors in (1) and (2), and denote $\boldsymbol{\beta} = (\alpha, \beta)'$ and $\boldsymbol{\phi} = (\mu, \phi)'$ as the vectors of unknown parameters in these equations; the respective OLS estimators are denoted as $\widehat{u}$, $\widehat{v}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\phi}}$ in the sequel. The following are useful results to prove the theoretical statements presented in the main text.

Lemma A1. *Under Assumptions 1 and 2, with either Assumption 3 or 3' it follows that,*
i) $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \beta_{uv} \left(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi} \right) + (X'X)^{-1}(X'\varepsilon)$.
ii) $\widehat{\boldsymbol{\phi}} = \boldsymbol{\phi} + (X'X)^{-1}(X'v)$.
iii) $v = \widehat{v} + X \left(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi} \right)$.
Proof. The proof of *i)* follows as in Stambaugh (1999). The proof of the remaining statements is immediate and follows from the standard OLS properties of equation (2). ■

Lemma A2. *Let* $\zeta_t = (\varepsilon_t, v_t)'$, *with* $\varepsilon_t = u_t - \beta_{uv} v_t$, *and define* $\psi_t = \tilde{x}_{t-1} \zeta_t$, $\sigma_x^2 = \sigma_v^2 / \left(1 - \phi^2\right)$ *and* $\Omega_\zeta = diag\left\{\sigma_\varepsilon^2, \sigma_v^2\right\}$, *with* $\sigma_\varepsilon^2 = \sigma_u^2 \left(1 - \rho^2\right)$. *Under Assumptions 1 and 2,* $\sum_{t=1}^{[Tr]} \zeta_t / \sqrt{T} \Rightarrow \Omega_\zeta^{1/2} \left(W_1(r), W_2(r)\right)' = \mathbf{B}(r)$, *in* $0 \leq r \leq 1$, *where* $\left(W_1(r), W_2(r)\right)'$ *is a two dimensional vector of independent Wiener processes in* $\mathcal{D}[0,1] \times \mathcal{D}[0,1]$. *Then, the following statements hold true under Assumptions 1 and 2, and as the sample size is allowed to diverge:*
i) Under Assumption 3, $T^{-1} \sum_{t=2}^{T} \tilde{x}_{t-1}^2 \xrightarrow{p} \sigma_x^2$, $T^{-1} \sum_{t=2}^{T} \widehat{v}_t^2 \xrightarrow{p} \sigma_v^2$, $T^{-1} \sum_{t=2}^{T} \tilde{x}_{t-1} v_t \xrightarrow{p} 0$, *and* $T^{-1/2} \sum_{t=2}^{T} \psi_t \Rightarrow \sigma_x \mathbf{B}(1)$;
ii) Under Assumption 3', $T^{-2} \sum_{t=2}^{T} \tilde{x}_{t-1}^2 \Rightarrow \sigma_v^2 \int_0^1 \overline{J}_c^2(r) dr$, $T^{-1} \sum_{t=2}^{T} \widehat{v}_t^2 \xrightarrow{p} \sigma_v^2$, *and* $T^{-1} \sum_{t=2}^{T} \psi_t \Rightarrow \sigma_v \int_0^1 \overline{J}_c(r) d\mathbf{B}(r)$, *with* $\overline{J}_c(r)$ *defined in Theorem 2.3, noting that* $W(r) \equiv W_2(r)$.

Proof. Under Assumption 2, the innovation vector ζ_t is a Gaussian i.i.d. process with zero mean and covariance matrix Ω_ζ bounded away from infinity and zero. Therefore, the invariance principle $T^{-1/2} \sum_{t=1}^{[Tr]} \zeta_t \Rightarrow \Omega_\zeta^{1/2} \mathbf{B}(r)$ holds; see Davidson (1994). For part *i)*, note that $\{x_t\}$ is a strictly stationary and ergodic process with finite second-order moments, while $\left\{\tilde{x}_{t-1}\varepsilon_t, \mathcal{F}^\zeta\right\}$ is a strictly stationary, ergodic and square-integrable MDS. The required results hold from the weak law of large numbers (WLLN) and the Central Limit Theorem; see White (2001) and Davidson (1994). Part *ii)* follows from the Continuous Mapping Theorem (CMP) and the convergence of partial sums to stochastic integrals; see Phillips (1987, 1988). ■

Proof of Theorem 2.1. Note that,

$$\left(\widehat{\beta}_F - \beta\right) = \frac{\sum_{t=2}^{T} \tilde{x}_{t-1} \varepsilon_t}{\sum_{t=2}^{T} \tilde{x}_{t-1}^2} + \beta_{uv} \left(\widehat{\phi} - \phi\right). \tag{A.7}$$

Hence, under Assumptions 1, 2 and 3, it follows that

$$\begin{aligned} \sqrt{T}\left(\widehat{\beta}_F - \beta\right) &= \frac{T^{-1/2} \sum_{t=2}^{T} \tilde{x}_{t-1} \varepsilon_t}{T^{-1} \sum_{t=2}^{T} \tilde{x}_{t-1}^2} + \beta_{uv} \frac{T^{-1/2} \sum_{t=2}^{T} \tilde{x}_{t-1} v_t}{T^{-1} \sum_{t=2}^{T} \tilde{x}_{t-1}^2} \\ &\Rightarrow \mathcal{N}\left(0, \frac{\sigma_\varepsilon^2 \left(1 - \phi^2\right)}{\sigma_v^2}\right) + \beta_{uv} \mathcal{N}(0, 1 - \phi^2) \end{aligned} \tag{A.8}$$

from Lemma A2*i*). Since the t-ratio is defined as

$$t_{\widehat{\beta}_F} = \sqrt{T}\left(\widehat{\beta}_F - \beta\right) \times \left(T^{-1}\sum_{t=2}^{T}\tilde{x}_{t-1}^2\right)^{1/2}\widehat{\sigma}_{\varepsilon}^{-1} \tag{A.9}$$

and $\widehat{\sigma}_{\varepsilon}^2 = T^{-1}\sum_{t=2}^{T}\varepsilon_t^2 + o_p(1)$, with $T^{-1}\sum_{t=2}^{T}\varepsilon_t^2 \xrightarrow{p} \sigma_{\varepsilon}^2$ from the WLLN, it follows directly from Cramér's Theorem and under the null hypothesis $\mathrm{H}_0 : \beta = 0$ that

$$t_{\widehat{\beta}_F} \Rightarrow \mathcal{Z} + \frac{\rho}{\sqrt{1-\rho^2}}\mathcal{Z} = \mathcal{N}\left(0, \frac{1}{1-\rho^2}\right) \tag{A.10}$$

as required. Alternatively, under Assumptions 1, 2 and 3', and operating in a similar way, we can show from Lemma A2*ii*) that

$$\begin{aligned} T\left(\widehat{\beta}_F - \beta\right) &= \frac{T^{-1}\sum_{t=2}^{T}\tilde{x}_{t-1}\varepsilon_t}{T^{-2}\sum_{t=2}^{T}\tilde{x}_{t-1}^2} + \frac{\beta_{uv}T^{-1}\sum_{t=2}^{T}\tilde{x}_{t-1}v_t}{T^{-2}\sum_{t=2}^{T}\tilde{x}_{t-1}^2} \\ &\Rightarrow \frac{\sigma_u}{\sigma_v}\frac{\sqrt{(1-\rho^2)}\int_0^1 \overline{J}_c(r)dW_1(r)}{\int_0^1 \overline{J}_c^2(r)dr} + \frac{\sigma_{uv}}{\sigma_v^2}\frac{\int_0^1 \overline{J}_c(r)dW_2(r)}{\int_0^1 \overline{J}_c^2(r)dr} \\ &= \frac{\sigma_u}{\sigma_v}\left[\frac{\sqrt{(1-\rho^2)}\int_0^1 \overline{J}_c(r)dW_1(r)}{\int_0^1 \overline{J}_c^2(r)dr} + \rho\frac{\int_0^1 \overline{J}_c(r)dW_2(r)}{\int_0^1 \overline{J}_c^2(r)dr}\right]. \end{aligned} \tag{A.11}$$

Noting that $t_{\widehat{\beta}_F} = T\left(\widehat{\beta}_F - \beta\right) \times \left(T^{-2}\sum_{t=2}^{T}\tilde{x}_{t-1}^2\right)^{1/2}/\widehat{\sigma}_{\varepsilon}$, it follows readily under the null hypothesis,

$$t_{\widehat{\beta}_F} \Rightarrow \mathcal{Z} + \rho\left(1-\rho^2\right)^{-1/2}\left\{\int_0^1\left[\overline{J}_c(r)\right]^2 dr\right\}^{-1/2}\int_0^1 \overline{J}_c(r)dW_2(r). \tag{A.12}$$

The statement in the text holds by recalling that $W(r) \equiv W_2(r)$. ■

Proof of Theorem 3.1. The limiting results obtain directly from Theorem 2.1 and Cramér's theorem, noting that $J_{x,T}(m) \Rightarrow J_c(m)$ under Assumptions 1, 2 and 3' as in Park (1990) and Park and Choi (1988). ■

Proof of Theorem 3.2. The OLS estimates can still be written as

$$\widehat{\beta}_{F,p} = \frac{\sum_{t=2}^{T}\tilde{x}_{t-1}\left(\tilde{y}_t - \widehat{\beta}_{ue}\widehat{e}_t\right)}{\sum_{t=2}^{T}\tilde{x}_{t-1}^2} \tag{A.13}$$

with $\widehat{e}_t$ denoting the corresponding residuals from the autoregression of the predictive variable. We first discuss the asymptotic behavior of this test under Assumptions 1, 2' and 3. The process $\{x_t\}$ obeys a a strictly stationary and ergodic AR$(p+1)$ process, so $\widehat{e}_t = \tilde{x}_t - \sum_{j=1}^{p+1}\widehat{\delta}_j^*\tilde{x}_{t-j}$, where $\widehat{\delta}_j^* = \left(\widehat{\delta}_j - \widehat{\delta}_{j-1}\widehat{\phi}\right)$, $j = 1, ..., p$ for $\delta_0 = 1$ and $\delta_{p+1}^* \equiv \delta_p\phi$. Noting that

$$\widehat{e}_t = e_t - \left(\widehat{\mathbf{\Phi}} - \mathbf{\Phi}\right)'\tilde{X}_{t-1} \tag{A.14}$$

with $\mathbf{\Phi} = \left(\delta_1^*, ..., \delta_p^*\right)'$, $\tilde{X}_{t-1} = (\tilde{x}_{t-1}, ..., \tilde{x}_{t-p-1})'$, and after denoting as $\widehat{\mathbf{\Phi}}$ the OLS counterpart of $\mathbf{\Phi}$, we can write

$$\begin{aligned}
\sqrt{T}\left(\widehat{\beta}_{F,p} - \beta\right) &= \left[\frac{T^{-1/2}\sum_{t=2}^{T}\tilde{x}_{t-1}\varepsilon_t}{T^{-1}\sum_{t=2}^{T}\tilde{x}_{t-1}^2}\right] + \left[\widehat{\beta}_{ue}\frac{\left(T^{-1}\sum_{t=2}^{T}\tilde{x}_{t-1}\tilde{X}_{t-1}'\right)\left[\sqrt{T}\left(\widehat{\mathbf{\Phi}} - \mathbf{\Phi}\right)\right]}{T^{-1}\sum_{t=2}^{T}\tilde{x}_{t-1}^2}\right] \\
&= \left[I_{1T}\right] + \left[I_{2T}\right], \qquad \text{(A.15)}
\end{aligned}$$

with I_{1T}, and I_{2T} defined implicitly. Redefine the random innovations $\zeta = (\varepsilon_t, e_t)'$ and note that $\left\{\zeta_t, \mathcal{F}_t^{\zeta}\right\}$ is a stationary and ergodic MDS vector with finite variance, and so is $\left\{\tilde{x}_{t-1}\zeta_t, \mathcal{F}_t^{\zeta}\right\}$, since, under Assumption 2' and from the Cauchy-Schwartz inequality, $\mathrm{E}\left(\tilde{x}_{t-1}^2\zeta_{it}^2\right) \leq \mathrm{E}\left(\tilde{x}_{t-1}^4\right)\mathrm{E}\left(\zeta_{it}^4\right) < \infty$. Therefore, using the appropriate central limit theory (White 2001, cf. Theorem 5.25), we can show as in Lemma A2*i*) that $I_{1T} \Rightarrow \mathcal{N}(0, \sigma_a^2/\sigma_x^2)$.

In order to analyze the asymptotic convergence of I_{2T}, note that $\tilde{x}_{t-1}\tilde{X}_t = (\tilde{x}_{t-1}^2, ..., \tilde{x}_{t-1}\tilde{x}_{t-p-1})'$ is a stationary and ergodic vector, so $T^{-1}\sum_{t=2}^{T}\tilde{x}_{t-1}\tilde{X}_{t-1} \xrightarrow{p} \mathbf{\Gamma}_1$ holds from the ergodic theorem (Davidson 1994, cf. Theorem 13.2), where $\mathbf{\Gamma}_1 = \left(\sigma_x^2, \gamma_1, ..., \gamma_p\right)'$, with $\gamma_j = E\left(\tilde{x}_t\tilde{x}_{t-j}\right)$ denoting the j-th autocovariance. Next, note that $\left(\widehat{\mathbf{\Phi}} - \mathbf{\Phi}\right) = \left[\sum_{t=2}^{T}\left(\tilde{X}_{t-1}\tilde{X}_{t-1}'\right)\right]^{-1} \times \left[\sum_{t=2}^{T}\tilde{X}_{t-1}e_t\right]$, where $\left\{\tilde{X}_{t-1}e_t, \mathcal{F}_t^{\zeta}\right\}$ is a stationary and ergodic MDS vector. Therefore, invoking a suitable CLT (Davidson 1994, Theorem 13.2) together with the Cramér-Wold device, and noting that $\sum_{t=2}^{T}\left(\tilde{X}_{t-1}\tilde{X}_{t-1}'\right) \xrightarrow{p} \mathbf{\Gamma}$ from ergodicity, it follows readily that $\sqrt{T}\left(\widehat{\mathbf{\Phi}} - \mathbf{\Phi}\right) \Rightarrow \mathcal{N}(\mathbf{0}, \sigma_e^2\mathbf{\Gamma}^{-1})$, with

$$\mathbf{\Gamma} = \begin{bmatrix} \sigma_x^2 & \gamma_1 & \cdots & \gamma_p \\ \gamma_1 & \sigma_x^2 & \cdots & \gamma_{p-1} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_p & \gamma_{p-1} & \cdots & \sigma_x^2 \end{bmatrix}. \qquad \text{(A.16)}$$

Then, since $T^{-1}\sum_{t=2}^{T}\tilde{x}_{t-1}^2 \xrightarrow{p} \sigma_x^2$ from the WLLN, we have from Cramér's Theorem that

$$I_{2T} \Rightarrow \mathcal{N}(0, \frac{\sigma_{ue}^2}{\sigma_e^2\sigma_x^4}\mathbf{\Gamma}_1'\mathbf{\Gamma}^{-1}\mathbf{\Gamma}_1), \qquad \text{(A.17)}$$

but since $\mathbf{\Gamma}_1'\mathbf{\Gamma}^{-1}\mathbf{\Gamma}_1 = \sigma_x^2$ and, recalling that $\sigma_\varepsilon^2 = \sigma_u^2(1-\rho^2)$, we conclude after trivial algebra that $\sqrt{T}\left(\widehat{\beta}_{F,p} - \beta\right) \Rightarrow \mathcal{N}\left(0, \sigma_\varepsilon^2\left[\left(1-\rho^2\right)\sigma_x^2\right]^{-1}\right)$, which is a result equivalent to the one obtained in Theorem 3.2 under Assumptions 1-3. Thus, it is now immediate to show that

$$t_{\widehat{\beta}_{F,p}}^{*} \Rightarrow \mathcal{N}\left(0, \frac{1}{1-\rho^2}\right) \qquad \text{(A.18)}$$

as required.

Consider now the behavior of the statistic under Assumptions 1, 2' and 3'. Hence, the largest autoregressive root of x_t obeys near-integrated dynamics, whereas the remaining roots are inside the unit circle. In this case, it is convenient to consider alternatively an equivalent representation of the autoregressive process that generates $\tilde{x}_t$. Thus, following

Fuller (1996), we can write $\tilde{x}_t = \phi^* \tilde{x}_{t-1} + \sum_{k=1}^{p} \vartheta_k \Delta \tilde{x}_{t-k} + e_t$, with $\phi^* = \phi + (1-\phi) \sum_{j=1}^{p} \delta_j$, and

$$\vartheta_j = \begin{cases} -\delta_j \phi + (1-\phi) \sum_{k=j+1}^{p} \delta_k, & 1 \le j \le p-1 \\ -\delta_p \phi, & j = p \end{cases} . \tag{A.19}$$

Since $\widehat{e}_t = e_t - \left(\widehat{\phi} - \phi\right) \tilde{x}_{t-1} - \sum_{k=1}^{p} \left(\widehat{\vartheta}_k - \vartheta_k\right) \Delta \tilde{x}_{t-k}$, it follows that,

$$\begin{aligned} T\left(\widehat{\beta}_{F,p} - \beta\right) &= \left[\frac{T^{-1}\sum_{t=2}^{T} \tilde{x}_{t-1}\varepsilon_t}{T^{-2}\sum_{t=2}^{T} \tilde{x}_{t-1}^2}\right] + \left[\widehat{\beta}_{ue} T\left(\widehat{\phi}^* - \phi^*\right)\right] + \left[T \frac{\sum_{t=2}^{T} \tilde{x}_{t-1} \sum_{k=1}^{p} \left(\widehat{\vartheta}_k - \vartheta_k\right) \Delta \tilde{x}_{t-k}}{\sum_{t=2}^{T} \tilde{x}_{t-1}^2}\right] \\ &= [I_{3T}] + [I_{4T}] + [I_{5T}] \end{aligned} \tag{A.20}$$

with these terms defined implicitly.

Let $(W_1(r), W_2(r))'$ be a two dimensional vector standard Brownian motion in $\mathcal{D}[0,1] \times \mathcal{D}[0,1]$. As in Lemma A2, the partial sum process obeys $T^{-1/2} \sum_{t=2}^{[Tr]} \zeta_t \Rightarrow (\sigma_\varepsilon W_1(r), \sigma_e W_2(r))'$, and $T^{-1/2} x_{[Tr]} \Rightarrow \omega J_c(r)$, $0 \le r \le 1$, $\omega^2 = \sigma_e^2 / A^2(1)$. Therefore, it follows from the CMT and as in Phillips (1987, 1988) that

$$I_{3T} \Rightarrow \frac{\sigma_u \sqrt{(1-\rho^2)} \int_0^1 \overline{J}_c(r) dW_1(r)}{\omega \int_0^1 \overline{J}_c^2(r) dr} \tag{A.21}$$

and, similarly,

$$I_{4T} \Rightarrow \beta_{ue} \omega \sigma_e \int_0^1 \overline{J}_c(r) dW_2(r) \left(\omega^2 \int_0^1 \overline{J}_c^2(r) dr\right)^{-1} . \tag{A.22}$$

Finally,

$$I_{5T} = \sum_{k=1}^{p-1} \sqrt{T} \left(\widehat{\vartheta}_k - \vartheta_k\right) \left(\frac{T^{-3/2} \sum_{t=2}^{T} \tilde{x}_{t-1} \Delta \tilde{x}_{t-k}}{T^{-2} \sum_{t=2}^{T} \tilde{x}_{t-1}^2}\right) = O_p\left(T^{-1/2}\right) = o_p(1) \tag{A.23}$$

because $\widehat{\vartheta}_k$ is a $\sqrt{T}-$consistent estimate of ϑ_k and, $T^{-1} \sum_{t=2}^{T} \tilde{x}_{t-1} \Delta \tilde{x}_{t-k} \Rightarrow \omega^2 \sigma_e^2 \int_0^1 \overline{J}_c(r) dW_2(r)$. Consequently,

$$T\left(\widehat{\beta}_{F,p} - \beta\right) \Rightarrow \frac{\sigma_u}{\omega_e} \left[\frac{\sqrt{(1-\rho^2)} \int_0^1 \overline{J}_c(r) dW_1(r)}{\int_0^1 \overline{J}_c^2(r) dr} + \rho \frac{\int_0^1 \overline{J}_c(r) dW_2(r)}{\int_0^1 \overline{J}_c^2(r) dr}\right] \tag{A.24}$$

and, therefore,

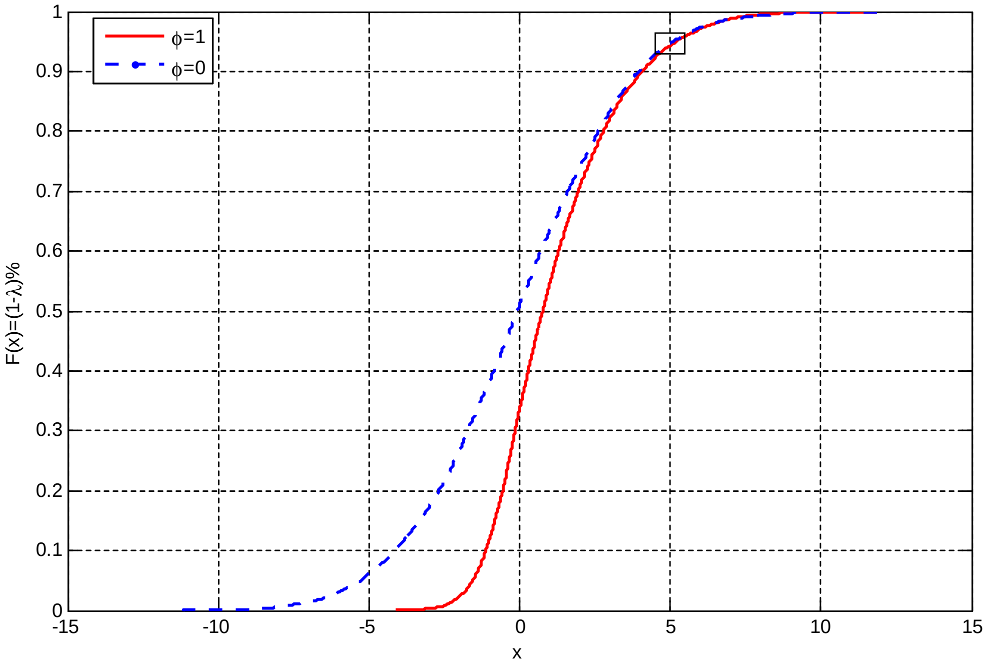
$$\begin{aligned} t^*_{\beta_{F,p}} &\Rightarrow \frac{\int_0^1 \overline{J}_c(r) dW_1(r)}{\left(\int_0^1 \overline{J}_c^2(r) dr\right)^{1/2}} + \frac{\rho}{\sqrt{(1-\rho^2)}} \frac{\int_0^1 \overline{J}_c(r) dW_2(r)}{\left(\int_0^1 \overline{J}_c^2(r) dr\right)^{1/2}} \\ &= \mathcal{Z} + \frac{\rho}{\sqrt{(1-\rho^2)}} \left(\int_0^1 \overline{J}_c^2(r) dr\right)^{-1/2} \int_0^1 \overline{J}_c(r) dW_2(r) \end{aligned} \tag{A.25}$$

which is asymptotically equivalent to the distribution of $t_{\widehat{\beta}_F}$ under Assumptions 1, 2 and 3'. ■

Proof of Theorem 3.4. The limiting results obtain directly from Theorem 3.3 and Cramér's theorem, noting that $J_{x,T}(m) \Rightarrow J_c(m)$ when replacing Assumption 2 with 2'. This completes the proof. ■

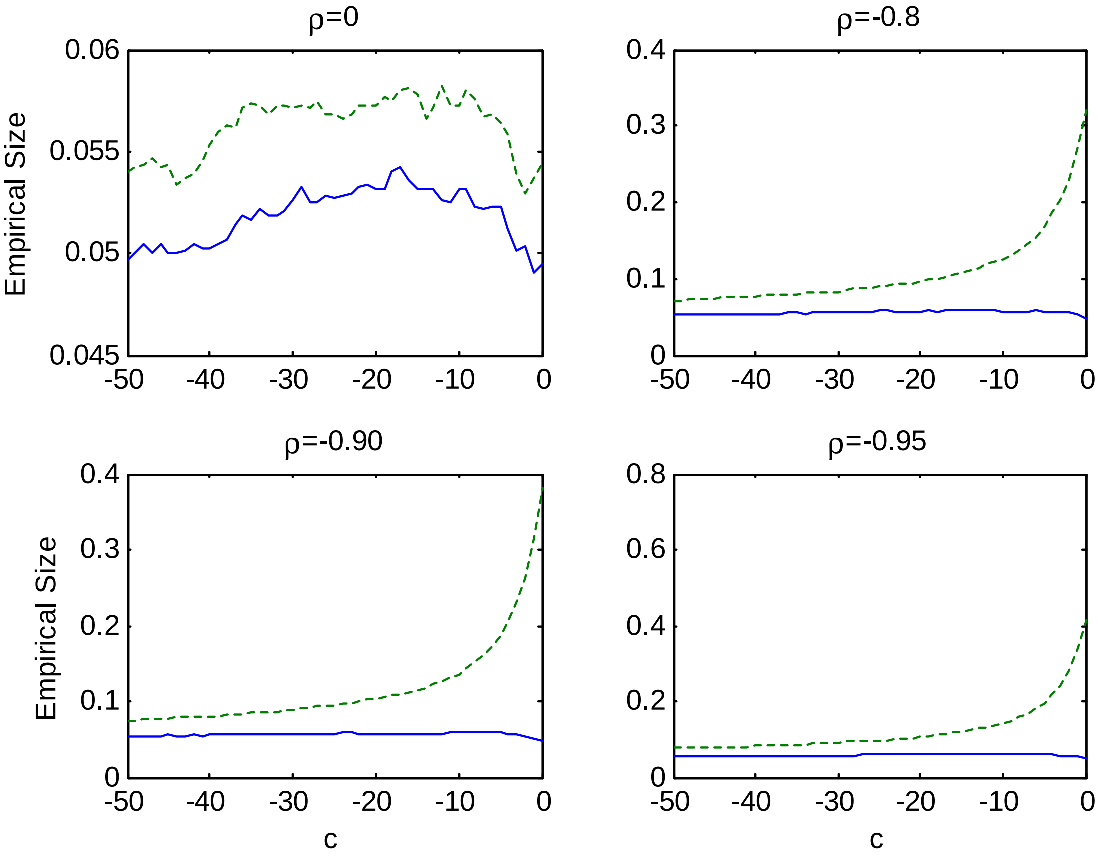
Figures

Figure 3.1: Cumulative distribution functions of $t^*_{\widehat{\beta},\lambda}$



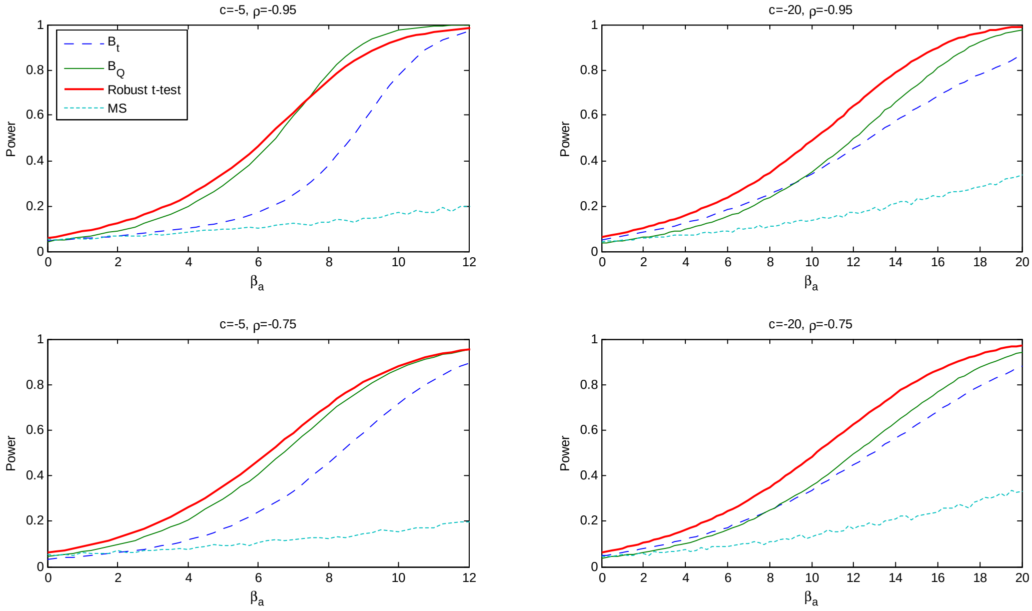
Note: Asymptotic cumulative distribution functions of the modified t-test $t_{\widehat{\beta},\lambda}$ for $\phi = 1$ (solid line) and $\phi = 0$ (dashed line) given $\rho = -0.95$ and $b^*_{\rho,\lambda}$ computed for a 95% right-tailed test.

Figure 4.1: Empirical sizes for T=100.



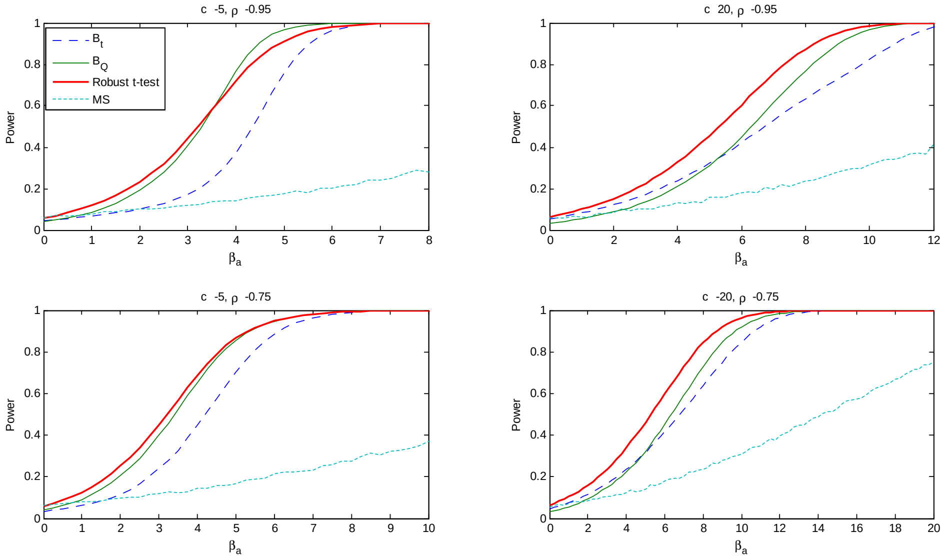
Note: Empirical sizes of the modified t-test $t_{\widehat{\beta},\lambda}$ (solid line) from an augmented predictive regression and the standard t-test (dashed line) from a standard predictive regression. Both tests are one-sided and considered at the 5% nominal level for the value of ρ displayed, $T = 100$.

Figure 4.2: Power functions based on i.i.d. innovations



Note: Power against local alternatives of the form $\beta = \beta_a/T$, with β_a in the range $[0, 20]$ given different values of $(c, \rho)'$ and i.i.d. innovations. The tests considered in the analysis are the Bonferroni t-test, B_t, the Bonferroni Q test, B_Q, the robust t-test $t_{\widehat{\beta},\lambda}$, and the non-parametric test, MS.

Figure 4.3: Power functions based on dependent innovations



Note: Power against local alternatives of the form $\beta = \beta_a/T$, with β_a in the range $[0, 20]$ given different values of $(c, \rho)'$ and innovations following an AR(1) process with coefficient $\delta = 0.5$. The tests considered in the analysis are the Bonferroni t-test, B_t, the Bonferroni Q test, B_Q, the robust t-test $t_{\widehat{\beta},\lambda}$, and the non-parametric test, MS.

Table 4.1: Critical values for the modified $t^*_{\widehat{\beta},\lambda}\left(\theta\right)$ tests

		0.0100	0.0500	0.1000	0.5000	0.9000	0.9500	0.9750	0.9900
0	T=100	-2.3492	-1.6613	-1.2867	0.0096	1.3108	1.7172	2.0504	2.3878
	T=250	-2.3047	-1.6495	-1.2750	0.0270	1.2750	1.6339	1.9226	2.2949
	T=500	-2.3116	-1.6077	-1.2534	0.0148	1.2898	1.6295	1.9461	2.3028
	T=1000	-2.2940	-1.6190	-1.2529	-0.0022	1.2799	1.6677	1.9683	2.3358
-0.1	T=100	-2.4025	-1.7074	-1.3202	-0.0176	1.2766	1.6763	2.0162	2.3739
	T=250	-2.3577	-1.6918	-1.3049	-0.0032	1.2536	1.6130	1.9080	2.2816
	T=500	-2.3462	-1.6383	-1.2639	-0.0090	1.2788	1.6237	1.9167	2.2877
	T=1000	-2.3318	-1.6543	-1.2846	-0.0204	1.2745	1.6660	1.9627	2.3301
-0.2	T=100	-2.4705	-1.7773	-1.3684	-0.0526	1.2596	1.6552	2.0062	2.3754
	T=250	-2.4260	-1.7485	-1.3679	-0.0258	1.2483	1.6029	1.9214	2.2779
	T=500	-2.4214	-1.6838	-1.3079	-0.0309	1.2758	1.6170	1.8995	2.2834
	T=1000	-2.3789	-1.6978	-1.3267	-0.0419	1.2691	1.6494	1.9676	2.3499
-0.3	T=100	-2.5698	-1.8635	-1.4462	-0.0840	1.2757	1.6576	2.0122	2.4112
	T=250	-2.5001	-1.8257	-1.4253	-0.0644	1.2518	1.6252	1.9593	2.2861
	T=500	-2.4936	-1.7486	-1.3641	-0.0573	1.2881	1.6331	1.9374	2.2895
	T=1000	-2.4322	-1.7648	-1.3743	-0.0673	1.2840	1.6866	2.0158	2.3899
-0.4	T=100	-2.7063	-1.9626	-1.5307	-0.1181	1.2925	1.6750	2.0466	2.4843
	T=250	-2.6240	-1.9461	-1.5117	-0.1082	1.2818	1.6647	2.0206	2.3586
	T=500	-2.6275	-1.8439	-1.4514	-0.0861	1.3082	1.6839	1.9971	2.3275
	T=1000	-2.5688	-1.8640	-1.4537	-0.0928	1.3091	1.7194	2.0635	2.4681
-0.5	T=100	-2.8999	-2.0849	-1.6611	-0.1683	1.3432	1.7549	2.1143	2.5717
	T=250	-2.8267	-2.0987	-1.6265	-0.1355	1.3333	1.7370	2.1034	2.4888
	T=500	-2.7819	-1.9965	-1.5744	-0.1179	1.3542	1.7754	2.1047	2.4885
	T=1000	-2.7493	-2.0038	-1.5703	-0.1228	1.3710	1.7978	2.1530	2.5681
-0.6	T=100	-3.1689	-2.2951	-1.8445	-0.2154	1.4037	1.8458	2.2361	2.6642
	T=250	-3.1166	-2.2739	-1.8009	-0.1899	1.4096	1.8581	2.2335	2.7058
	T=500	-3.0686	-2.1965	-1.7460	-0.1611	1.4112	1.8795	2.2385	2.6865
	T=1000	-2.9920	-2.1787	-1.7360	-0.1525	1.4554	1.9234	2.3328	2.7799
-0.7	T=100	-3.6204	-2.6225	-2.1210	-0.2865	1.5303	2.0334	2.4600	2.9241
	T=250	-3.5468	-2.5640	-2.0521	-0.2558	1.5510	2.0640	2.4677	3.0013
	T=500	-3.4882	-2.4948	-1.9979	-0.2152	1.5292	2.0438	2.5113	2.9941
	T=1000	-3.3853	-2.4421	-1.9536	-0.1964	1.5951	2.1033	2.5830	3.1104
-0.8	T=100	-4.3167	-3.1545	-2.5397	-0.3901	1.7525	2.3570	2.8747	3.4153
	T=250	-4.3164	-3.0815	-2.4792	-0.3412	1.8031	2.4141	2.8782	3.4421
	T=500	-4.1578	-3.0098	-2.4167	-0.3060	1.7808	2.4017	2.9254	3.5953
	T=1000	-4.0821	-2.9404	-2.3503	-0.2681	1.8623	2.5217	3.0309	3.6845
-0.9	T=100	-5.8311	-4.3048	-3.4740	-0.5794	2.2676	3.1066	3.7234	4.5195
	T=250	-5.8782	-4.2185	-3.3869	-0.5080	2.3766	3.1842	3.8641	4.7141
	T=500	-5.7596	-4.1876	-3.3827	-0.4675	2.3618	3.2384	3.9309	4.8453
	T=1000	-5.6735	-4.0893	-3.2778	-0.4216	2.5172	3.3948	4.1356	5.0216
-0.95	T=100	-7.7399	-5.7147	-4.6256	-0.8185	2.9004	4.0315	4.9292	5.9929
	T=250	-8.0211	-5.7684	-4.6546	-0.7279	3.1554	4.2295	5.1694	6.3454
	T=500	-7.9215	-5.7844	-4.6761	-0.6953	3.2659	4.3705	5.3933	6.5447
	T=1000	-7.8824	-5.7237	-4.5584	-0.6092	3.4714	4.6876	5.7366	6.9599

Table 4.2: Asymptotic b-values for the t-test

	$b^*_{\lambda,\rho}$		
ρ	10%	5%	1%
0.00	0.0030	0.0000	0.0000
-0.10	0.0390	0.0192	0.0157
-0.20	0.8750	0.0657	0.0250
-0.30	0.1302	0.0950	0.0690
-0.40	0.1980	0.1670	0.1096
-0.50	0.2580	0.2120	0.1950
-0.60	0.3410	0.2580	0.2405
-0.70	0.4408	0.3620	0.3150
-0.80	0.5495	0.4270	0.3530
-0.90	0.6448	0.5870	0.4550
-0.95	0.7200	0.6250	0.5400

Note: Values of the equalizing constant $b^*_{\lambda,\rho}$ for the $\lambda\%$ one-sided test given different values of ρ.

Table 5.1: Predictability analysis

Panel A

	T	x_t	p	$\widehat{\beta}$	$\widehat{\beta}_{F,p}$	$t^*_{\widehat{\beta}_{F,p}}(\lambda)$	$\widehat{\rho}$
S&P500	123	d-p	3	0.057	0.061	2.402	-0.85
	123	e-p	1	0.091	-0.022	-2.579	-0.74
CRSP$_{Annual}$	77	d-p	1	0.158	0.126	2.325*	-0.70
	77	e-p	1	0.162	0.107	4.610*	-0.94
CRSP$_{Quarterly}$	305	d-p	1	0.034	0.021	2.799	-0.93
	305	e-p	1	0.047	0.030	8.490*	-0.98
CRSP$_{Monthly}$	913	d-p	2	0.008	0.004	2.015	-0.95
	913	e-p	1	0.013	0.007	7.606*	-0.99

Panel B

	T	x_t	p	$\widehat{\beta}$	$\widehat{\beta}_{F,p}$	$t^*_{\widehat{\beta}_{F,p}}(\lambda)$	$\widehat{\rho}$
S&P500	115	d-p	3	0.083	0.088	2.586	-0.84
	115	e-p	1	0.125	-0.028	-0.881	-0.69
CRSP$_{Annual}$	69	d-p	1	0.269	0.22	3.038*	-0.67
	69	e-p	1	0.269	0.186	6.746*	-0.94
CRSP$_{Quarterly}$	273	d-p	1	0.054	0.034	3.639	-0.93
	273	e-p	1	0.077	0.052	12.478*	-0.98
CRSP$_{Monthly}$	817	d-p	2	0.012	0.006	2.488	-0.94
	817	e-p	2	0.021	0.013	9.646*	-0.98

Panel C

	T	x_t	p	$\widehat{\beta}$	$\widehat{\beta}_{F,p}$	$t^*_{\widehat{\beta},\lambda,p}(\theta)$	$\widehat{\rho}$
CRSP$_{Annual}$	51	d-p	1	0.154	0.113	1.463	-0.72
	51	e-p	1	0.113	0.073	1.697	-0.94
	51	r3	2	-1	-1.051	-1.162	0.05
	51	y-r1	1	1.847	1.527	0.943	-0.24
CRSP$_{Quarterly}$	204	d-p	1	0.038	0.024	3.036	-0.97
	204	e-p	1	0.029	0.018	2.589	-0.98
	204	r3	4	-0.374	-0.392	-1.847	-0.09
	204	y-r1	2	0.84	0.816	1.937*	-0.10
CRSP$_{Monthly}$	612	d-p	1	0.012	0.007	2.805	-0.96
	612	e-p	1	0.009	0.005	2.304	-0.98
	612	r3	2	-0.15	-0.157	-2.49	-0.07
	612	y-r1	1	0.366	0.362	2.926*	-0.07

Note: In Panel A, the S&P500 sample covers the period from 1880 to 1994 whereas the CRSP sample starts in 1926 and ends in 1994. In Panel B, the S&P500 sample covers the period from 1880 to 1994 whereas the CRSP sample starts in 1926 and ends in 1994. In Panel C, only CRSP data is considered and the sample starts in 1952 and ends in 2002. In the table, T denotes the sample length; $\widehat{\beta}$ and $\widehat{\beta}_{F,p}$ denote, respectively, the OLS slope estimate in the non-augmented predictive regression and the augmented predictive regression with $p+1$ lags; $t^*_{\widehat{\beta}_{F,p}}(\lambda)$ is the value of robust t-test and $\widehat{\rho}$ is the estimated correlation coefficient in the empirical innovations. The predictors are the dividend-price ratio (d-p), the earnings-price ratio (e-p), the three-month T-bill rate (r3) and the long-short yield spread

WORKING PAPERS

2010

1/10 MEASURING COMOVEMENT IN THE TIME-FREQUENCY SPACE
— *António Rua*

2/10 EXPORTS, IMPORTS AND WAGES: EVIDENCE FROM MATCHED FIRM-WORKER-PRODUCT PANELS
— *Pedro S. Martins, Luca David Opromolla*

3/10 NONSTATIONARY EXTREMES AND THE US BUSINESS CYCLE
— *Miguel de Carvalho, K. Feridun Turkman, António Rua*

4/10 EXPECTATIONS-DRIVEN CYCLES IN THE HOUSING MARKET
— *Luisa Lambertini, Caterina Mendicino, Maria Teresa Punzi*

5/10 COUNTERFACTUAL ANALYSIS OF BANK MERGERS
— *Pedro P. Barros, Diana Bonfim, Moshe Kim, Nuno C. Martins*

6/10 THE EAGLE. A MODEL FOR POLICY ANALYSIS OF MACROECONOMIC INTERDEPENDENCE IN THE EURO AREA
— *S. Gomes, P. Jacquinot, M. Pisani*

7/10 A WAVELET APPROACH FOR FACTOR-AUGMENTED FORECASTING
— *António Rua*

8/10 EXTREMAL DEPENDENCE IN INTERNATIONAL OUTPUT GROWTH: TALES FROM THE TAILS
— *Miguel de Carvalho, António Rua*

9/10 TRACKING THE US BUSINESS CYCLE WITH A SINGULAR SPECTRUM ANALYSIS
— *Miguel de Carvalho, Paulo C. Rodrigues, António Rua*

10/10 A MULTIPLE CRITERIA FRAMEWORK TO EVALUATE BANK BRANCH POTENTIAL ATTRACTIVENESS
— *Fernando A. F. Ferreira, Ronald W. Spahr, Sérgio P. Santos, Paulo M. M. Rodrigues*

11/10 THE EFFECTS OF ADDITIVE OUTLIERS AND MEASUREMENT ERRORS WHEN TESTING FOR STRUCTURAL BREAKS IN VARIANCE
— *Paulo M. M. Rodrigues, Antonio Rubia*

12/10 CALENDAR EFFECTS IN DAILY ATM WITHDRAWALS
— *Paulo Soares Esteves, Paulo M. M. Rodrigues*

13/10 MARGINAL DISTRIBUTIONS OF RANDOM VECTORS GENERATED BY AFFINE TRANSFORMATIONS OF INDEPENDENT TWO-PIECE NORMAL VARIABLES
— *Maximiano Pinheiro*

14/10 MONETARY POLICY EFFECTS: EVIDENCE FROM THE PORTUGUESE FLOW OF FUNDS
— *Isabel Marques Gameiro, João Sousa*

15/10 SHORT AND LONG INTEREST RATE TARGETS
— *Bernardino Adão, Isabel Correia, Pedro Teles*

16/10 FISCAL STIMULUS IN A SMALL EURO AREA ECONOMY
— *Vanda Almeida, Gabriela Castro, Ricardo Mourinho Félix, José Francisco Maria*

17/10 FISCAL INSTITUTIONS AND PUBLIC SPENDING VOLATILITY IN EUROPE
— *Bruno Albuquerque*

18/10 GLOBAL POLICY AT THE ZERO LOWER BOUND IN A LARGE-SCALE DSGE MODEL
— *S. Gomes, P. Jacquinot, R. Mestre, J. Sousa*

19/10 LABOR IMMOBILITY AND THE TRANSMISSION MECHANISM OF MONETARY POLICY IN A MONETARY UNION
— *Bernardino Adão, Isabel Correia*

20/10 TAXATION AND GLOBALIZATION
— *Isabel Correia*

21/10 TIME-VARYING FISCAL POLICY IN THE U.S.
— *Manuel Coutinho Pereira, Artur Silva Lopes*

22/10 DETERMINANTS OF SOVEREIGN BOND YIELD SPREADS IN THE EURO AREA IN THE CONTEXT OF THE ECONOMIC AND FINANCIAL CRISIS
— *Luciana Barbosa, Sónia Costa*

23/10 FISCAL STIMULUS AND EXIT STRATEGIES IN A SMALL EURO AREA ECONOMY
— *Vanda Almeida, Gabriela Castro, Ricardo Mourinho Félix, José Francisco Maria*

24/10 FORECASTING INFLATION (AND THE BUSINESS CYCLE?) WITH MONETARY AGGREGATES
— *João Valle e Azevedo, Ana Pereira*

25/10 THE SOURCES OF WAGE VARIATION: AN ANALYSIS USING MATCHED EMPLOYER-EMPLOYEE DATA
— *Sónia Torres,Pedro Portugal, John T.Addison, Paulo Guimarães*

26/10 THE RESERVATION WAGE UNEMPLOYMENT DURATION NEXUS
— *John T. Addison, José A. F. Machado, Pedro Portugal*

27/10 BORROWING PATTERNS, BANKRUPTCY AND VOLUNTARY LIQUIDATION
— *José Mata, António Antunes, Pedro Portugal*

28/10 THE INSTABILITY OF JOINT VENTURES: LEARNING FROM OTHERS OR LEARNING TO WORK WITH OTHERS
— *José Mata, Pedro Portugal*

29/10 THE HIDDEN SIDE OF TEMPORARY EMPLOYMENT: FIXED-TERM CONTRACTS AS A SCREENING DEVICE
— *Pedro Portugal, José Varejão*

30/10 TESTING FOR PERSISTENCE CHANGE IN FRACTIONALLY INTEGRATED MODELS: AN APPLICATION TO WORLD INFLATION RATES
— *Luis F. Martins, Paulo M. M. Rodrigues*

31/10 EMPLOYMENT AND WAGES OF IMMIGRANTS IN PORTUGAL
— *Sónia Cabral, Cláudia Duarte*

32/10 EVALUATING THE STRENGTH OF IDENTIFICATION IN DSGE MODELS. AN A PRIORI APPROACH
— *Nikolay Iskrev*

33/10 JOBLESSNESS
— *José A. F. Machado, Pedro Portugal, Pedro S. Raposo*

2011

1/11 WHAT HAPPENS AFTER DEFAULT? STYLIZED FACTS ON ACCESS TO CREDIT
— *Diana Bonfim, Daniel A. Dias, Christine Richmond*

2/11 IS THE WORLD SPINNING FASTER? ASSESSING THE DYNAMICS OF EXPORT SPECIALIZATION
— *João Amador*

3/11 UNCONVENTIONAL FISCAL POLICY AT THE ZERO BOUND
— *Isabel Correia, Emmanuel Farhi, Juan Pablo Nicolini, Pedro Teles*

4/11 MANAGERS' MOBILITY, TRADE STATUS, AND WAGES
— *Giordano Mion, Luca David Opromolla*

5/11 FISCAL CONSOLIDATION IN A SMALL EURO AREA ECONOMY
— *Vanda Almeida, Gabriela Castro, Ricardo Mourinho Félix, José Francisco Maria*

6/11 CHOOSING BETWEEN TIME AND STATE DEPENDENCE: MICRO EVIDENCE ON FIRMS' PRICE-REVIEWING STRATEGIES
— *Daniel A. Dias, Carlos Robalo Marques, Fernando Martins*

7/11 WHY ARE SOME PRICES STICKIER THAN OTHERS? FIRM-DATA EVIDENCE ON PRICE ADJUSTMENT LAGS
— *Daniel A. Dias, Carlos Robalo Marques, Fernando Martins, J. M. C. Santos Silva*

8/11 LEANING AGAINST BOOM-BUST CYCLES IN CREDIT AND HOUSING PRICES
— *Luisa Lambertini, Caterina Mendicino, Maria Teresa Punzi*

9/11 PRICE AND WAGE SETTING IN PORTUGAL LEARNING BY ASKING
— *Fernando Martins*

10/11 ENERGY CONTENT IN MANUFACTURING EXPORTS: A CROSS-COUNTRY ANALYSIS
— *João Amador*

11/11 ASSESSING MONETARY POLICY IN THE EURO AREA: A FACTOR-AUGMENTED VAR APPROACH
— *Rita Soares*

12/11 DETERMINANTS OF THE EONIA SPREAD AND THE FINANCIAL CRISIS
— *Carla Soares, Paulo M. M. Rodrigues*

13/11 STRUCTURAL REFORMS AND MACROECONOMIC PERFORMANCE IN THE EURO AREA COUNTRIES: A MODEL-BASED ASSESSMENT
— *S. Gomes, P. Jacquinot, M. Mohr, M. Pisani*

14/11 RATIONAL VS. PROFESSIONAL FORECASTS
— *João Valle e Azevedo, João Tovar Jalles*

15/11 ON THE AMPLIFICATION ROLE OF COLLATERAL CONSTRAINTS
— *Caterina Mendicino*

16/11 MOMENT CONDITIONS MODEL AVERAGING WITH AN APPLICATION TO A FORWARD-LOOKING MONETARY POLICY REACTION FUNCTION
— *Luis F. Martins*

17/11 BANKS' CORPORATE CONTROL AND RELATIONSHIP LENDING: EVIDENCE FROM RETAIL LOANS
— *Paula Antão, Miguel A. Ferreira, Ana Lacerda*

18/11 MONEY IS AN EXPERIENCE GOOD: COMPETITION AND TRUST IN THE PRIVATE PROVISION OF MONEY
— *Ramon Marimon, Juan Pablo Nicolini, Pedro Teles*

19/11 ASSET RETURNS UNDER MODEL UNCERTAINTY: EVIDENCE FROM THE EURO AREA, THE U.K. AND THE U.S.
— *João Sousa, Ricardo M. Sousa*

20/11 INTERNATIONAL ORGANISATIONS' *VS*. PRIVATE ANALYSTS' FORECASTS: AN EVALUATION
— *Ildeberta Abreu*

21/11 HOUSING MARKET DYNAMICS: ANY NEWS?
— *Sandra Gomes, Caterina Mendicino*

22/11 MONEY GROWTH AND INFLATION IN THE EURO AREA: A TIME-FREQUENCY VIEW
— *António Rua*

23/11 WHY EX(IM)PORTERS PAY MORE: EVIDENCE FROM MATCHED FIRM-WORKER PANELS
— *Pedro S. Martins, Luca David Opromolla*

24/11 THE IMPACT OF PERSISTENT CYCLES ON ZERO FREQUENCY UNIT ROOT TESTS
— *Tomás del Barrio Castro, Paulo M.M. Rodrigues, A.M. Robert Taylor*

25/11 THE TIP OF THE ICEBERG: A QUANTITATIVE FRAMEWORK FOR ESTIMATING TRADE COSTS
— *Alfonso Irarrazabal, Andreas Moxnes, Luca David Opromolla*

26/11 A CLASS OF ROBUST TESTS IN AUGMENTED PREDICTIVE REGRESSIONS
— *Paulo M.M. Rodrigues, Antonio Rubia*