

## A Bootstrap Stationarity Test for Predictive Regression Invalidity

Iliyan Georgiev, David I. Harvey, Stephen J. Leybourne & A. M. Robert Taylor

To cite this article: Iliyan Georgiev, David I. Harvey, Stephen J. Leybourne & A. M. Robert Taylor (2019) A Bootstrap Stationarity Test for Predictive Regression Invalidity, Journal of Business & Economic Statistics, 37:3, 528-541, DOI: [10.1080/07350015.2017.1385467](https://doi.org/10.1080/07350015.2017.1385467)

To link to this article: <https://doi.org/10.1080/07350015.2017.1385467>



© 2019 The Author(s). Published with license by Taylor & Francis Group, LLC© Iliyan Georgiev, David I. Harvey, Stephen J. Leybourne, and A. M. Robert Taylor.



[View supplementary material](#)



Published online: 31 May 2018.



[Submit your article to this journal](#)



Article views: 1705



[View related articles](#)



[View Crossmark data](#)



Citing articles: 4 [View citing articles](#)

# A Bootstrap Stationarity Test for Predictive Regression Invalidity

Iliyan GEORGIEV

Department of Economics, University of Bologna, 40126 Bologna BO, Italy ([i.georgiev@unibo.it](mailto:i.georgiev@unibo.it))

David I. HARVEY and Stephen J. LEYBOURNE

School of Economics, University of Nottingham, Nottingham NG7 2RD, United Kingdom  
([dave.harvey@nottingham.ac.uk](mailto:dave.harvey@nottingham.ac.uk); [steve.leybourne@nottingham.ac.uk](mailto:steve.leybourne@nottingham.ac.uk))

A. M. Robert TAYLOR 

Essex Business School, University of Essex, Colchester CO4 3SQ, United Kingdom ([rtaylor@essex.ac.uk](mailto:rtaylor@essex.ac.uk))

In order for predictive regression tests to deliver asymptotically valid inference, account has to be taken of the degree of persistence of the predictors under test. There is also a maintained assumption that any predictability in the variable of interest is purely attributable to the predictors under test. Violation of this assumption by the omission of relevant persistent predictors renders the predictive regression invalid, and potentially also spurious, as both the finite sample and asymptotic size of the predictability tests can be significantly inflated. In response, we propose a predictive regression invalidity test based on a stationarity testing approach. To allow for an unknown degree of persistence in the putative predictors, and for heteroscedasticity in the data, we implement our proposed test using a fixed regressor wild bootstrap procedure. We demonstrate the asymptotic validity of the proposed bootstrap test by proving that the limit distribution of the bootstrap statistic, conditional on the data, is the same as the limit null distribution of the statistic computed on the original data, conditional on the predictor. This corrects a long-standing error in the bootstrap literature whereby it is incorrectly argued that for strongly persistent regressors and test statistics akin to ours the validity of the fixed regressor bootstrap obtains through equivalence to an unconditional limit distribution. Our bootstrap results are therefore of interest in their own right and are likely to have applications beyond the present context. An illustration is given by reexamining the results relating to U.S. stock returns data in Campbell and Yogo (2006). Supplementary materials for this article are available online.

KEY WORDS: Conditional distribution; Fixed regressor wild bootstrap; Granger causality; Persistence; Predictive regression; Stationarity test.

## 1. INTRODUCTION

Predictive regression (hereafter PR) is a widely used tool in applied finance and economics, and forms the basis for Granger causality testing. A very common application is in the context of testing the linear rational expectations hypothesis. A core example of this is testing whether future (excess) stock returns are predictable (Granger caused) by current information, such as the dividend yield or the term structure of interest rates. Often it is found that the posited predictor variable (e.g., dividend yield) exhibits persistence behavior akin to a (near) unit root autoregressive process, while the variable being predicted (e.g., the stock return) resembles a (near) martingale difference sequence (m.d.s.).

In basic form, a test of predictability involves running an OLS regression of the variable being predicted,  $y_t$  say, on the lagged value of a posited predictor variable,  $x_t$  say, and testing the significance of the estimated coefficient on  $x_{t-1}$  using a standard regression  $t$ -ratio. Here, the null hypothesis is that  $y_t$  is unpredictable (in mean) from ex ante information; the alternative is that it is predictable from  $x_{t-1}$ . Cavanagh, Elliott, and Stock (CES; 1995) showed that when the innovation driving  $x_t$  is correlated with  $y_t$  (as is often thought to be case in practice,

for example, the stock price is a component of both the return and the dividend yield), then these tests can be badly over-sized if  $x_t$  is a local to unit root process but critical values appropriate for the case where  $x_t$  is a pure unit root process are used. This over-size can be interpreted as a tendency toward finding spurious predictability in  $y_t$ , in that it is incorrectly concluded that  $x_{t-1}$  can be used to predict  $y_t$  when in fact  $y_t$  is unpredictable; see also Rossi (2005) for a discussion of related issues. Attempting to address this issue, CES discuss Bonferroni bound-based procedures that yield conservative tests, while Campbell and Yogo (CY; 2006) considered a point optimal variant of the  $t$ -test and employed confidence belts. Phillips (2014) proposed a

© 2019 Iliyan Georgiev, David I. Harvey, Stephen J. Leybourne, and A. M. Robert Taylor.

This is an Open Access article. Non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly attributed, cited, and is not altered, transformed, or built upon in any way, is permitted. The moral rights of the named author(s) have been asserted.

Journal of Business & Economic Statistics  
July 2019, Vol. 37, No. 3

DOI: 10.1080/07350015.2017.1385467

modification to the test proposed in CY, which is asymptotically valid in the case where  $x_t$  can be either local-to-unity or stationary. Recently, Breitung and Demetrescu (BD; 2015) considered variable addition and instrumental variable (IV) methods to correct test size. Near-optimal PR tests can also be found in Elliott, Müller, and Watson (2015) and Jansson and Moreira (2006).

A misspecified PR of  $y_t$  on  $x_{t-1}$  (with nonzero slope) can also arise from these tests in cases where  $y_t$  is in fact predictable and is Granger-caused (possibly by the process  $\{x_t\}$  and) by some other persistent process,  $\{z_t\}$  say. The variable  $z_t$  might be a manifest variable or an unobserved latent variable. (We distinguish between Granger causality, defined by conditioning on counterfactual information sets that can be chosen to contain the past of the variable  $z$ , observable or not, and predictability as a pragmatic concept based on available observations. Where  $z_t$  is latent it cannot therefore be termed a predictor.) Here, and in the special case where  $x_{t-1}$  is an invalid predictor variable (because  $y_t$  is Granger-caused solely by  $\{z_t\}$  and  $x_t$  is uncorrelated with  $z_t$ ), it is known that the regression of  $y_t$  on  $x_{t-1}$  can lead to serious upward size distortions in the standard PR tests, with the same conclusion of spurious predictability of  $y_t$  by  $x_{t-1}$  as discussed earlier; see Ferson, Sarkissian, and Simin (2003a, b) and Deng (2014). More generally, where both  $\{x_t\}$  and  $\{z_t\}$  Granger-cause  $y_t$ , or  $x_t$  and  $z_t$  are correlated, a linear predictor of  $y_t$  by  $x_{t-1}$  would still be misspecified because it would be suboptimal with respect to quadratic loss, even if the optimal linear predictor based on observables might involve  $x_{t-1}$ . (Even where  $y_t$  is not Granger-caused by  $\{x_t\}$  but  $z_t$  is a latent variable correlated with  $x_t$ ,  $x_{t-1}$  would pick up some of the information from the past of  $z_t$  and so  $x_{t-1}$  would not be a spurious predictor variable.) Specifically, in this case the optimal linear predictor for  $y_t$  would involve the past of  $z_t$  (if  $z_t$  is a manifest variable), or further variables among the lags of both  $y_t$  and  $x_{t-1}$  (if  $z_t$  is latent). This fundamental misspecification problem in the estimated PR will affect all of the predictability tests discussed above.

We demonstrate theoretically and by means of simulations the potential for a misspecified PR of  $y_t$  on  $x_{t-1}$  to arise in the context of a model where  $x_t$  and  $z_t$  follow persistent processes, which we model as local-to-unity autoregressions, while modeling the coefficient on  $z_{t-1}$  as being local-to-zero. As a consequence, it is important to be able to identify, a priori, if  $y_t$  is Granger caused by some ignored  $\{z_t\}$ . Our approach involves testing for persistence in the residuals from a regression of  $y_t$  on  $x_{t-1}$ . Consequently, any effect that  $x_{t-1}$  may have on  $y_t$ , through the value of its slope coefficient in the putative PR, is eliminated from the residuals, and any persistence they display thereafter is attributable to the unincluded variable  $z_{t-1}$ , and would signal that the PR is misspecified. The test for PR misspecification we suggest is based on the co-integration tests of Shin (1994) and Leybourne and McCabe (1994), themselves variants of the stationarity test of Kwiatkowski et al. (KPSS; 1992). Although originally designed to detect pure unit root behavior in regression residuals, Müller (2005) showed that these tests also reject when near unit root behavior is present, making them well-suited to the testing scenario of this article.

An issue arising with our proposed test is that under its null hypothesis that  $z_{t-1}$  plays no role in the data-generating process [DGP] for  $y_t$ , its limit distribution depends on the local-to-unity parameter in the process for  $x_t$ , even though

the residuals used are invariant to the coefficient on  $x_{t-1}$  in the DGP. In principle, this makes it difficult to control the size of the test. However, we show a bootstrap procedure which treats  $x_{t-1}$  as a fixed regressor (i.e., the observed  $x_{t-1}$  is used in calculating bootstrap analogs of our test statistic) can be implemented to yield an asymptotically size-controlled test. This *fixed regressor bootstrap* approach is not itself new to the literature and has been employed by, among others, Gonçalves and Kilian (2004) and Hansen (2000). Because many financial and economic time series are thought to display nonstationary volatility and/or conditional heteroscedasticity in their innovations, it is also important for our proposed testing procedure to be (asymptotically) robust to these effects. We therefore use a heteroscedasticity-robust variant of the fixed regressor bootstrap along the lines proposed by Hansen (2000). This uses a wild bootstrap scheme to generate bootstrap analogs of  $y_t$ . We show that our proposed fixed regressor wild bootstrap test has local asymptotic power against the same local alternatives that give rise to a misspecified PR of  $y_t$  on  $x_{t-1}$ .

We establish large-sample validity of our bootstrap method by showing that the limit distribution of the bootstrap statistic, conditional on the data, is the same as the limit null distribution of the statistic computed on the original data, conditional on the posited predictor variable. Our method of proof has wider applicability to other scenarios where a fixed regressor bootstrap is used with (near-) integrated regressors. For instance, our proof corrects an error in the bootstrap literature arising from Hansen (2000) who incorrectly suggested, in the context of a closely related test statistic, that for strongly persistent regressors the validity of the fixed regressor bootstrap is due to the coincidence of the unconditional null limit distribution of the original statistic with that of the limit distribution of the bootstrap statistic conditional of the data; actually, by following our proof, this coincidence can be seen not to occur for Hansen's statistic.

The article is organized as follows. Section 2 presents the maintained DGP and sets out the various null and alternative hypotheses regarding predictability of  $y_t$  by  $x_{t-1}$  and  $z_{t-1}$ . To aid lucidity, we consider a single putative predictor variable,  $x_t$ , and single unincluded variable,  $z_t$ , both with m.d.s. errors. Generalizations to richer model specifications are straightforward and discussed at various points. Section 3 details the asymptotic distributions of standard PR statistics under the various hypotheses, demonstrating the inference problems caused by unincluded persistent variables. Section 4 introduces our proposed test for PR invalidity, detailing its limit distribution and showing the validity of the fixed regressor wild bootstrap scheme in providing asymptotic size control. The asymptotic power of this procedure is also examined here and compared with the degree of size distortions associated with PR tests. Section 5 presents the results of a set of finite sample simulations investigating the size and power of our proposed bootstrap tests. An empirical illustration reconsidering the results pertaining to U.S stock returns data in CY is given in Section 6. Proofs and additional simulation results appear in a supplementary appendix.

We use the following notation:  $\lfloor \cdot \rfloor$  is the floor function;  $\mathbb{I}(\cdot)$  is the indicator function;  $x := y$  ( $x =: y$ ) means that  $x$  is defined by  $y$  ( $y$  is defined by  $x$ );  $\xrightarrow{w}$  and  $\xrightarrow{D}$  for weak convergence and convergence in probability, respectively. For a vector,  $x$ ,  $\|x\| := (x'x)^{1/2}$ , the Euclidean norm. Finally,  $D^k := D_k[0, 1]$  is the space of right continuous with left limit (càdlàg) functions

from  $[0, 1]$  to  $\mathbb{R}^k$ , equipped with the Skorokhod topology, and  $\mathcal{D} := \mathcal{D}^1$ .

## 2. THE MODEL AND PREDICTABILITY HYPOTHESES

The basic DGP we consider for observed  $y_t$  is

$$y_t = \alpha_y + \beta_x x_{t-1} + \beta_z z_{t-1} + \epsilon_{yt}, \quad t = 1, \dots, T, \quad (1)$$

where  $x_t$  and  $z_t$  satisfy

$$x_t = \alpha_x + s_{x,t}, \quad z_t = \alpha_z + s_{z,t}, \quad t = 0, \dots, T \quad (2)$$

$$s_{x,t} = \rho_x s_{x,t-1} + \epsilon_{xt}, \quad s_{z,t} = \rho_z s_{z,t-1} + \epsilon_{zt}, \quad t = 1, \dots, T, \quad (3)$$

where  $\rho_x := 1 - c_x T^{-1}$  and  $\rho_z := 1 - c_z T^{-1}$ , with  $c_x \geq 0$  and  $c_z \geq 0$ , so that  $x_t$  and  $z_t$  are unit root or local-to-unit root autoregressive processes. We let  $s_{x,0}$  and  $s_{z,0}$  be  $O_p(1)$  variates. Following CES and to examine the asymptotic local power of the test procedures we discuss, we parameterize  $\beta_x$  and  $\beta_z$  as  $\beta_x = g_x T^{-1}$  and  $\beta_z = g_z T^{-1}$ , respectively, which entails that when  $g_x$  and/or  $g_z$  are nonzero,  $y_t$  is a persistent, but local-to-noise process. (Notice that an observationally equivalent formulation of the model can be obtained by treating  $\beta_x$  and  $\beta_z$  as fixed constants but parameterizing the variances of  $\epsilon_{xt}$  and  $\epsilon_{zt}$  to be local-to-zero; see, in particular, the discussion following Equation (10) later. We choose the local-to-zero coefficient formulation for consistency with CES.)

Our interest lies in examining the behavior of predictability tests derived from the PR of  $y_t$  on  $x_{t-1}$  when  $y_t$  is generated by the DGP in (1)–(3) with  $\beta_z \neq 0$ , and subsequently developing tests for the null hypothesis that  $\beta_z = 0$ . In doing so, it is important to note that the motivating issue of spurious predictability of  $y_t$  by  $x_{t-1}$ , in the case where there is no correlation between  $x_{t-1}$  and  $z_{t-1}$ , arises whenever  $x_{t-1}$  and the unobserved  $z_{t-1}$  are both persistent processes. In the general case where no dependence restrictions are placed between  $x_{t-1}$  and  $z_{t-1}$ , the presence of  $z_{t-1}$  in (1) does not entail that  $x_{t-1}$  is a spurious predictor for  $y_t$ . Rather it implies that the PR of  $y_t$  on  $x_{t-1}$  alone is misspecified.

In the context of (1),  $z_{t-1}$  could be either an omitted manifest variable or an unobserved latent variable. An example of the latter is given by the case where  $y_t$  are (currency, commodity, or bond) returns and  $x_{t-1}$  is either the lagged forward premium (spot minus forward price/rate) or a lagged futures basis (spot minus futures price/rate). Here, there is an unobserved latent risk premium which is believed to be strongly persistent, and which in combination with the strongly persistent predictor has been suggested as a possible driver for empirically unorthodox findings, such as the well-known forward premium (or Fama) puzzle; see Gospodinov (2009). A second example is provided by the long-run risk model of Bansal and Yaron (2004). Certain versions of their model can be rewritten as PRs for returns with an unobserved long-run persistent component in consumption. In the latent case, it would also be quite reasonable to view  $z_t$  not through a literal interpretation of the DGP in (1)–(3) but rather as a general proxy for underlying misspecification in the PR, under which interpretation it would clearly not make sense for  $z_t$  to be stationary rather than persistent. Possible examples are provided by the case where the coefficient on  $x_{t-1}$  displays time-varying behavior, such as has been considered in, for example, Paye and Timmermann (2006) and Cai, Wang, and Wang (2015), or where

the data on  $x_t$  are observed with a strongly persistent measurement error driven by relatively low variance innovations.

The innovation vector  $\epsilon_t := [\epsilon_{xt}, \epsilon_{zt}, \epsilon_{yt}]'$  is taken to satisfy the following conditions:

*Assumption 1.* The innovation process  $\epsilon_t$  can be written as  $\epsilon_t = HD_t e_t$  where:

(a)  $H$  and  $D_t$  are the  $3 \times 3$  nonstochastic matrices

$$H := \begin{bmatrix} h_{11} & 0 & 0 \\ h_{21} & h_{22} & 0 \\ h_{31} & h_{32} & h_{33} \end{bmatrix}, \quad D_t := \begin{bmatrix} d_{1t} & 0 & 0 \\ 0 & d_{2t} & 0 \\ 0 & 0 & d_{3t} \end{bmatrix}$$

with  $h_{ij} \in \mathbb{R}$ ,  $h_{ii} > 0$  ( $i, j = 1, 2, 3$ ), and  $HH'$  strictly positive definite. The volatility terms  $d_{it}$  satisfy  $d_{it} = d_i(t/T)$ , where  $d_i \in \mathcal{D}$  are nonstochastic, strictly positive functions.

(b)  $e_t$  is a  $3 \times 1$  vector martingale difference sequence (m.d.s.) with respect to a filtration  $\mathcal{F}_t$ , to which it is adapted, with conditional covariance matrix  $\sigma_t := E(e_t e_t' | \mathcal{F}_{t-1})$  satisfying: (i)  $T^{-1} \sum_{t=1}^T \sigma_t \xrightarrow{p} E(e_t e_t') = I_3$ ; (ii)  $\sup_t E \|e_t\|^{4+\delta} < \infty$  for some  $\delta > 0$ .

*Remark 1.* Assumption 1 implies that  $\epsilon_t$  is a vector m.d.s. relative to  $\mathcal{F}_t$ , with conditional variance matrix  $\Omega_{t|t-1} := E(\epsilon_t \epsilon_t' | \mathcal{F}_{t-1}) = (HD_t) \sigma_t (HD_t)'$ , and time-varying unconditional variance matrix  $\Omega_t := E(\epsilon_t \epsilon_t') = (HD_t)(HD_t)'$ . Stationary conditional heteroscedasticity and nonstationary unconditional volatility are obtained as special cases with  $D_t = I_3$  (constant unconditional variance, hence only conditional heteroscedasticity) and  $\sigma_t = I_3$  (so  $\Omega_{t|t-1} = \Omega_t = \Omega(t/T)$ , only unconditional nonstationary volatility), respectively. (The assumption that  $E(e_t e_t') = I_3$  made in part (b) (i) and the parameterization of the unconditionally homoscedastic case by  $D_t = I_3$  are without loss of generality, by nonidentification considerations.) As discussed in Cavaliere, Rahbek, and Taylor (2010), Assumption 1(a) implies that the elements of  $\Omega_t$  are only required to be bounded and to display a countable number of jumps, therefore allowing for an extremely wide class of potential models for the behavior of the variance matrix of  $\epsilon_t$ , including single or multiple variance or covariance shifts, variances which follow a broken trend, and smooth transition variance shifts.

*Remark 2.* Under Assumption 1, an identification issue regarding the parameters  $\beta_x$ ,  $\beta_z$ , and  $h_{21}$  arises in the case where  $c_x = c_z$ . In this case, whenever the observables  $(y_t, x_t)$  satisfy (1) for certain  $\beta_x, \beta_z \neq 0$ , and  $z_t$ , they also satisfy (1) for  $\beta_x^\lambda = \beta_x + \lambda$ ,  $\beta_z^\lambda = \beta_z$ , and  $z_t^\lambda = z_t - \lambda \beta_z^{-1} x_t$ , for any  $\lambda$ , where  $z_t^\lambda$  is a (local-to-) unit root autoregressive process and its innovations  $\epsilon_{zt}^\lambda = \epsilon_{zt} - \lambda \beta_z^{-1} \epsilon_{xt}$  are such that  $[\epsilon_{xt}, \epsilon_{zt}^\lambda, \epsilon_{yt}]'$  satisfies Assumption 1, upon a redefinition of the matrix  $H$ . In particular, if  $\beta_z \neq 0$ , then it is possible to choose  $\lambda = h_{21} h_{11}^{-1} \beta_z$  such that  $\epsilon_{xt}$  and  $\epsilon_{zt}^\lambda$ , the innovations driving  $x_t$  and  $z_t^\lambda$ , respectively, are uncorrelated. In accordance with OLS identification conditions, we will discuss the predictive implications of (1) under the identifying condition  $E(\epsilon_{xt} \epsilon_{zt}) = 0$  (equivalently,  $h_{21} = 0$ ) if  $\beta_z \neq 0$ , and under the condition  $\beta_z = 0$  otherwise. In the case where  $z_t$  is a named latent variable (such as an unobserved risk



premium) or a manifest variable, the value of  $E(\epsilon_{xt}\epsilon_{zt})$  is implicitly fixed by the choice of  $z_t$  and an alternative is to discuss (1) by using this value for identification.

*Remark 3.* We notice that a PR based on  $x_{t-1}$  alone is misspecified whenever  $\beta_z \neq 0$ , regardless of the value of either  $\beta_x$  or the correlation between  $\epsilon_{xt}$  and  $\epsilon_{zt}$ . If  $h_{21} = 0$ ,  $x_{t-1}$ , and  $z_{t-1}$  would be uncorrelated with one another and any conclusion of predictability from the PR of  $y_t$  on  $x_{t-1}$  in the case where  $\beta_x = 0$  and  $\beta_z \neq 0$  in (1) would be purely spurious because the best linear predictor (BLP; with respect to symmetric quadratic loss) of  $y_t$  given the past of  $\{y_t, x_t\}$  would not involve  $x_{t-1}$ , although the BLP with respect to a larger information set might involve  $x_{t-1}$ . When  $h_{21} \neq 0$ ,  $x_{t-1}$ , and  $z_{t-1}$  are correlated, and thus, for forecasting purposes,  $x_{t-1}$  could act as a proxy for the information in  $z_{t-1}$ . Nonetheless, if  $\beta_z \neq 0$ , the BLP of  $y_t$  would not be a function of  $x_{t-1}$  alone: for a manifest variable  $z_t$ , the BLP given the past of  $\{y_t, x_t, z_t\}$  would involve  $z_{t-1}$ , whereas for a latent variable  $z_t$ , the BLP given the past of  $\{y_t, x_t\}$  would involve lags of  $y_t$  and  $x_t$  (even if  $\beta_x = 0$ , as some of the predictive power of  $z_{t-1}$  would be picked up by  $x_{t-1}$ ).

*Remark 4.* For transparency, the structure in (1)–(3) is exposited for a scalar variable,  $z_t$ . This is without loss of generality, as one may consider that  $z_t = \gamma'z_t^*$  where  $z_t^*$  is a vector of variables, which might therefore contain both omitted manifest and latent variables.

We are now ready to discuss, in the context of (1), the possibilities for the predictability and causation of  $y_t$  by the variables  $x_{t-1}$  and  $z_{t-1}$ , focusing on linear predictors. One potential case that has received much attention in the literature is that where  $y_t$  is Granger-caused only by the process  $\{x_t\}$ , so that it is predictable only by  $x_{t-1}$ , implying that  $\beta_x \neq 0$  while  $\beta_z = 0$  in (1). This forms the alternative hypothesis in the PR tests discussed in Section 3, where the corresponding null is that  $\beta_x = 0$ , and, in the context of our model, the maintained hypothesis that  $\beta_z = 0$ , so that  $y_t$  is unpredictable under the null. However, it is also a possibility that  $y_t$  is Granger-caused only by the process  $\{z_t\}$ , unincorporated in the PR. In this case,  $\beta_x = 0$  and  $\beta_z \neq 0$ , thereby violating the aforementioned maintained hypothesis, and a PR of  $y_t$  on  $x_{t-1}$  alone would be misspecified, regardless of whether  $z_t$  is a manifest or latent variable (see Remark 3). In the special case where  $h_{21} = 0$  and  $x_{t-1}$  does not enter the BLP of  $y_t$ , a conclusion to the contrary is an instance of spurious predictability. A final possibility is that  $\beta_x \neq 0$  and  $\beta_z \neq 0$  so that  $y_t$  is Granger-caused by both processes  $\{x_t\}$  and  $\{z_t\}$ . In this last case if  $z_t$  was an omitted manifest variable then a correctly specified PR could be obtained by including  $z_{t-1}$  in the PR. If, on the other hand,  $z_t$  was a latent variable, a correctly specified BLP of  $y_t$  would include more observables (e.g.,  $y_{t-1}$ ) than  $x_{t-1}$ . We summarize these four cases using the following taxonomy of hypotheses within the context of DGP (1):

- $H_u : \beta_x = 0, \beta_z = 0$   $y_t$  is unpredictable (in mean)
- $H_x : \beta_x \neq 0, \beta_z = 0$   $y_t$  is Granger-caused by  $\{x_t\}$  alone
- $H_z : \beta_x = 0, \beta_z \neq 0$   $y_t$  is Granger-caused by  $\{z_t\}$  alone
- $H_{xz} : \beta_x \neq 0, \beta_z \neq 0$   $y_t$  is Granger-caused by  $\{x_t\}$  and  $\{z_t\}$ .

In hypothesis testing terms, standard PR tests attempt to distinguish between the null  $H_u$  and the alternative  $H_x$ . Here, we consider the impact of the presence of  $z_{t-1}$  in the DGP on such tests, that is, we investigate the behavior of PR tests of  $H_u$  against  $H_x$  when in fact  $H_z$  or  $H_{xz}$  is true. In addition, we propose a test for possible PR invalidity, where the appropriate composite null is  $H_u$  or  $H_x$  ( $H_u, H_x$ ), and the alternative  $H_z$  or  $H_{xz}$  ( $H_z, H_{xz}$ ).

We end this section by stating some implications of Assumption 1 for our asymptotic analysis. Associated with a standard Brownian motion  $B = [B_1, B_2, B_3]'$  in  $\mathbb{R}^3$ , let  $B_\eta = [B_{\eta_1}, B_{\eta_2}, B_{\eta_3}]'$  be the heteroscedastic Gaussian motion defined by  $B_{\eta_i}(r) := f_i^{-1/2} \int_0^r d_i(s)dB_i(s)$ ,  $r \in [0, 1]$ , where  $f_i := \int_0^1 d_i(s)^2 ds$ ,  $i = 1, 2, 3$ . We can also write  $B_{\eta_i} \stackrel{d}{=} B_i(\eta_i)$ ,  $i = 1, 2, 3$ , where  $\eta_i$  denotes the variance profile  $\eta_i(r) := f_i^{-1} \int_0^r d_i(s)^2 ds$ ,  $r \in [0, 1]$ , such that  $B_{\eta_i}$  is a time-changed Brownian motion; see, for example, Davidson (1994, p. 486). In particular,  $\eta_i(r) = r$ ,  $r \in [0, 1]$ , under unconditional homoscedasticity. Then the following functional weak convergence result holds in  $\mathcal{D}^3 \times \mathbb{R}^{3 \times 3}$ , by Lemma 1 of Boswijk et al. (2016):

$$\left( T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \epsilon_t, T^{-1} \sum_{t=1}^T \sum_{s=1}^{t-1} \epsilon_s \epsilon_t' \right) \xrightarrow{w} \left( M_\eta(r), \int_0^1 M_\eta(s) dM_\eta(s)' \right), \quad r \in [0, 1], \quad (4)$$

where  $M_\eta := [M_{\eta_x}, M_{\eta_z}, M_{\eta_y}]' := HF^{1/2}B_\eta$  for the diagonal matrix  $F := \text{diag}\{f_1, f_2, f_3\}$ . Let  $\Omega_\eta := \{\omega_{ab}\}_{a,b \in \{x,y,z\}} := \text{var}\{M_\eta(1)\} = HFH'$ , which in the unconditionally homoscedastic case  $D_t = I_3$  reduces to

$$\begin{aligned} HH' &= \begin{bmatrix} h_{11}^2 & h_{11}h_{21} & h_{11}h_{31} \\ h_{11}h_{21} & h_{21}^2 + h_{22}^2 & h_{21}h_{31} + h_{22}h_{32} \\ h_{11}h_{31} & h_{21}h_{31} + h_{22}h_{32} & h_{31}^2 + h_{32}^2 + h_{33}^2 \end{bmatrix} \\ &=: \begin{bmatrix} \sigma_{xx} & \sigma_{xz} & \sigma_{xy} \\ \sigma_{xz} & \sigma_{zz} & \sigma_{zy} \\ \sigma_{xy} & \sigma_{zy} & \sigma_{yy} \end{bmatrix} =: \Omega. \end{aligned}$$

It will prove convenient to define the two Ornstein–Uhlenbeck-type processes  $M_{\eta_c,u}(r) := \int_0^r e^{(s-r)c_u} dM_{\eta_u}(s)$  for  $u = x, z$  and  $r \in [0, 1]$ , along with the standardized analogs  $B_{\eta_c,u}(r) := \omega_{uu}^{-1/2} M_{\eta_c,u}(r)$  and their demeaned counterparts  $\tilde{B}_{\eta_c,u}(r) := B_{\eta_c,u}(r) - \int_0^1 B_{\eta_c,u}(s)$ .

### 3. ASYMPTOTIC BEHAVIOR OF PREDICTIVE REGRESSION TESTS

To fix ideas, as in CES, we first consider the basic PR test of  $H_u$  against  $H_x$ , based on the  $t$ -ratio for testing  $\beta_x = 0$  in the fitted linear regression

$$y_t = \hat{\alpha}_y + \hat{\beta}_x x_{t-1} + \hat{\epsilon}_{yt}, \quad t = 1, \dots, T. \quad (5)$$

The test statistic is given by

$$t_u := \frac{\hat{\beta}_x}{\sqrt{s_y^2 / \sum_{t=1}^T (x_{t-1} - \bar{x}_{-1})^2}}, \quad \hat{\beta}_x := \frac{\sum_{t=1}^T (x_{t-1} - \bar{x}_{-1})y_t}{\sum_{t=1}^T (x_{t-1} - \bar{x}_{-1})^2}$$

and  $s_y^2 := (T - 2)^{-1} \sum_{t=1}^T \hat{\epsilon}_{yt}^2$ , with  $\bar{x}_{-1} := T^{-1} \sum_{t=1}^T x_{t-1}$ .

In addition to the  $t$ -test, we also analyze a point optimal variant introduced by CY. For a known value of  $\rho_x$ , the (infeasible) test statistic takes the following form:

$$Q := \frac{\hat{\beta}_x - (s_{xy}/s_x^2)(\hat{\rho}_x - \rho_x)}{\sqrt{s_y^2\{1 - (s_{xy}^2/s_x^2 s_y^2)\} / \sum_{t=1}^T (x_{t-1} - \bar{x}_{-1})^2}}$$

where  $\hat{\beta}_x$  and  $s_y^2$  are as defined above,  $s_{xy} := (T - 2)^{-1} \sum_{t=1}^T \hat{\epsilon}_{xt}\hat{\epsilon}_{yt}$  and  $s_x^2 := (T - 2)^{-1} \sum_{t=1}^T \hat{\epsilon}_{xt}^2$  with  $\hat{\epsilon}_{xt}$  denoting the OLS residuals from regressing  $x_t$  on a constant and  $x_{t-1}$ , and where  $\hat{\rho}_x := \sum_{t=1}^T (x_{t-1} - \bar{x}_{-1})x_t / \sum_{t=1}^T (x_{t-1} - \bar{x}_{-1})^2$ . In the case where  $s_{xy} = 0$ ,  $Q$  and  $t_u$  coincide.

The limit distributions of  $t_u$  and  $Q$  under Assumption 1 are shown in the next theorem.

*Theorem 1.* For the DGP (1), (2), (3) and under Assumption 1, the weak limits of  $t_u$  and  $Q$  as  $T \rightarrow \infty$  are of the form

$$\frac{\int_0^1 \bar{M}_{\eta c,x}(r) dN_{\eta y}(r)}{\sqrt{\int_0^1 \bar{M}_{\eta c,x}(r)^2}} + \frac{g_x \int_0^1 \bar{M}_{\eta c,x}(r)^2 + g_z \int_0^1 \bar{M}_{\eta c,x}(r) M_{\eta c,z}(r)}{\sqrt{n_y \int_0^1 \bar{M}_{\eta c,x}(r)^2}}, \tag{6}$$

where  $\bar{M}_{\eta c,x}(r) := M_{\eta c,x}(r) - \int_0^1 M_{\eta c,x}(s) ds$ ,  $r \in [0, 1]$ , and  $N_{\eta y}, n_y$  are statistic-specific. Thus, for the  $t_u$  statistic,  $N_{\eta y} := \omega_{yy}^{-1/2} M_{\eta y}$  and  $n_y := \omega_{yy}$ , whereas for the  $Q$  statistic,  $N_{\eta y} := \omega_{y|x}^{-1/2} \{M_{\eta y} - \omega_{xy} \omega_{xx}^{-1} M_{\eta x}\}$  and  $n_y := \omega_{yy} - \omega_{xy}^2 / \omega_{xx} =: \omega_{y|x}$ .

*Remark 5.* Notice that the limit expressions for  $t_u$  and  $Q$  in (6) are identical when  $h_{31} = 0$  (i.e.,  $\omega_{xy} = 0$ ). The limit expression in (6) shows the dependence of  $t_u$  and  $Q$  on  $g_z$  under  $H_z$  (where  $g_x = 0$  but  $g_z \neq 0$ ). Consequently, even for infeasible versions of these tests where all other nuisance parameters were known, the use of asymptotic critical values appropriate for these tests under  $H_u$  will not result in size-controlled procedures under  $H_z$  and raises the possibility that spurious rejections in favor of predictability of  $y_t$  by  $x_{t-1}$  will be encountered when  $y_t$  is actually predictable by  $z_{t-1}$  (see Ferson Sarkissian, and Simin 2003a, Ferson Sarkissian, and Simin 2003a,b, and Deng 2014, for related results under nonlocalized  $\beta_z$ ). Under  $H_{xz}$ , where both  $g_x \neq 0$  and  $g_z \neq 0$ , any rejection by  $t_u$  or  $Q$  could not uniquely be ascribed to the role of  $x_{t-1}$ , potentially suggesting the existence of a well-specified PR that is in fact under-specified due to the omission of  $z_{t-1}$ . The same issues also hold for the feasible versions of the  $t_u$  and  $Q$  tests developed in CES and in CY and Phillips (2014), respectively.

*Remark 6.* In the special case where  $c_x = c_z$ , the limit of  $t_u$  in (6) can be written as

$$\frac{\int_0^1 \bar{B}_{\eta c,x}(r) dM_{\eta y}(r)}{\sqrt{\omega_{yy} \int_0^1 \bar{B}_{\eta c,x}(r)^2}} + g_x^{\perp} \left( \frac{\omega_{xx}}{\omega_{yy}} \right)^{1/2} \sqrt{\int_0^1 \bar{B}_{\eta c,x}(r)^2} + g_z \left( \frac{\omega_{z|x}}{\omega_{yy}} \right)^{1/2} \frac{\int_0^1 \bar{B}_{\eta c,x}(r) B_{\eta c,2}(r)}{\sqrt{\int_0^1 \bar{B}_{\eta c,x}(r)^2}} \tag{7}$$

with  $B_{\eta c,2}(r) := \int_0^r e^{(s-r)c_z} dB_{\eta 2}(s)$  for  $r \in [0, 1]$ ,  $\omega_{z|x} := \omega_{zz} - \omega_{xz}^2 / \omega_{xx}$ , and  $g_x^{\perp} T^{-1} := (g_x + \omega_{xz} \omega_{xx}^{-1} g_z) T^{-1}$  representing the coefficient of  $x_{t-1}$  in a redefinition of (1) where  $x_{t-1}$  is orthogonal to the unincluded persistent variable (see Remark 2 with

$\lambda = h_{21} h_{11}^{-1} \beta_z = \omega_{xz} \omega_{xx}^{-1} g_z T^{-1}$ ). Not surprisingly, therefore,  $t_u$  can be anticipated to have relatively low power to reject  $H_u$  in favor of  $H_{xz}$  when the contribution of  $x_{t-1}$  to the variability of  $y_t$  (as measured by  $|g_x^{\perp}| \omega_{xx}^{1/2} \omega_{yy}^{-1/2}$ ) is low, and also the contribution of  $z_{t-1}$  corrected for  $x_{t-1}$  (as measured by  $|g_z| \omega_{z|x}^{1/2} \omega_{yy}^{-1/2}$ ) is low. Additionally, the correlation between  $\bar{B}_{\eta c,x}$  and  $M_{\eta y}$  (for  $h_{31} \neq 0$ ) renders the leading term in (7) non-Gaussian, affecting both the size and the power of the test. These comments also apply to the limit of the  $Q$  statistic, except that the first term in (7) is then standard Gaussian.

We will now proceed to investigate the extent of the size distortions that occur in the  $t_u$  and  $Q$  tests when  $g_z \neq 0$ . Before doing so, it should be noted that other PR tests have been proposed in the literature, including the near-optimal tests of Elliott, Müller, and Watson (2015) and Jansson and Moreira (2006); see the useful recent summaries provided in BD and Cai, Wang, and Wang (2015). The issues we discuss in this article are pertinent irrespective of which particular PR test one uses, in cases where the putative and unincluded predictors are persistent. They are also relevant for the case where a putative PR contains multiple predictors.

### 3.1 Asymptotic Size of Predictive Regression Tests Under $H_z$

To obtain as transparent as possible a picture of the large-sample size properties of  $t_u$  and  $Q$  under  $H_z$ , we abstract from any role that nonstationary volatility plays by setting  $d_i = 1$ ,  $i = 1, 2, 3$ . We then simulate the limit distributions using 10,000 Monte Carlo replications, approximating the Brownian motion processes in the limiting functionals for (6) using independent  $N(0, 1)$  random variates, with the integrals approximated by normalized sums of 2000 steps. Critical values are obtained by setting  $g_x = g_z = 0$ ; for  $t_u$  these depend on  $c_x$  and also (it can be shown)  $h_{31}^2 / (h_{31}^2 + h_{32}^2 + h_{33}^2) = \sigma_{xy}^2 / \sigma_{xx} \sigma_{yy}$ , while for  $Q$ , these depend on  $c_x$  alone. These quantities are assumed known, so we are essentially analyzing the large-sample behavior of infeasible variants of  $t_u$  and  $Q$ . We graph nominal 0.10-level sizes of two-sided tests as functions of the parameter  $g_z = \{0, 2.5, 5.0, \dots, 50.0\}$  with  $g_x = 0$ . For  $c_x = c_z = c = \{0, 10\}$ , we set  $\sigma_{xx} = \sigma_{zz} = \sigma_{yy} = 1$ , and consider  $\sigma_{xy} = \sigma_{zy} = 0$  plus  $\sigma_{xy} = -0.70$  with  $\sigma_{zy} = \{0, -0.70, 0.70\}$  where  $\sigma_{xz} = 0$  throughout. Setting  $c_x = c_z$  is not a requirement here, but simply facilitates keeping  $x_t$  and  $z_t$  balanced in terms of their persistence properties.

The results of this size simulation exercise are shown in Figure 1. For  $c = 0$  we observe the sizes of  $t_u$  and  $Q$  growing monotonically from the baseline 0.10 level with increasing  $g_z$ , thereby giving rise to an ever-increasing likelihood of ascribing spurious predictive ability to  $x_{t-1}$ . Both tests' sizes are seen to exceed 0.85 for  $g_z = 50$ , while even a value of  $g_z$  as small as  $g_z = 12.5$  produces sizes in excess of 0.50. The size patterns for  $t_u$  and  $Q$  are also quite similar, which is as we would expect given that  $g_z$  impacts upon their limit distributions in a very similar way. Of course, when  $\sigma_{xy} = 0$ , the tests have identical limits, while for  $\sigma_{xy} = -0.7$ , there is a general tendency for  $Q$  to show slightly more pronounced over-sizing than  $t_u$  (possibly reflecting the relatively higher power that this test can achieve under

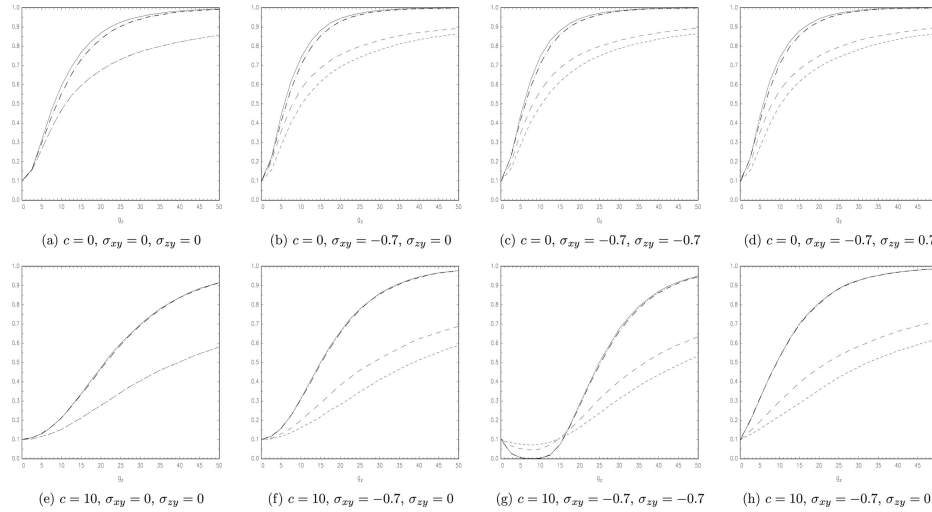


Figure 1. Asymptotic rejection frequencies of  $S$ ,  $S_B$  (power) and  $t_u$ ,  $Q$  (size):  $g_x = 0$ ,  $c_x = c_z = c$ ;  $S$ :  $\cdots$ ,  $S_B$ :  $—$ ,  $t_u$ :  $- - -$ ,  $Q$ :  $- \cdot -$ .

$H_x$ ). Size distortions appear little influenced by the value taken by  $\sigma_{zy}$ . With  $c = 10$  qualitatively, the same comments apply here as for the case  $c = 0$ . That said, we do observe that the over-sizing now manifests itself more slowly with increasing  $g_z$ . Indeed, when  $\sigma_{zy} = -0.70$  some modest under-size is observed for small values of  $g_z$ . However, both sizes are still above 0.50 once  $g_z = 50$  so spurious predictability does remain a serious issue. That the problem is less severe here simply reflects the fact that  $x_{t-1}$  and  $z_{t-1}$  are lower (but still high) persistence processes.

It would be difficult to argue that spurious predictive ability is not a potentially important consideration to take into account when employing either of the  $t_u$  and  $Q$  tests to infer predictability with high persistence processes. Although we have focused this analysis on OLS-based PR tests, similar qualitative results will pertain for other PR tests including the recently proposed IV-based tests of BD whenever a high persistence IV is used. A low persistence IV test should be less prone to over-size in the presence of a high persistence unincluded variable  $z_{t-1}$ , but the price paid for employing such an IV is that when a true predictor  $x_{t-1}$  is highly persistent, the IV test will have very poor power. Basically, whenever there is scope for high persistence properties of regressors to yield good power for PR tests, we should always remain alert to the possibility of spurious predictability.

#### 4. A TEST FOR PREDICTIVE REGRESSION INVALIDITY

Given the potential for standard PR tests to spuriously signal predictability of  $y_t$  by  $x_{t-1}$  (alone) when  $\beta_z \neq 0$ , we now consider a test devised to distinguish between  $\beta_z = 0$  and  $\beta_z \neq 0$ . Nonrejection by such a test would indicate that  $z_{t-1}$  plays no role in predicting  $y_t$ , and hence that standard PR tests based on  $x_{t-1}$  are valid. Rejection, however, would indicate the presence of an unincluded variable  $z_{t-1}$  in the DGP for  $y_t$ , signaling the invalidity of PR tests based on  $x_{t-1}$ . Formally, then, we wish to test the null hypothesis that  $\beta_z = 0$ , that is,  $H_u$ ,  $H_x$ , against the alternative that  $\beta_z \neq 0$ , that is,  $H_z$ ,  $H_{xz}$ , in (1).

#### 4.1 The Test Statistic and Conventional Asymptotics

The test we develop is based on testing a null hypothesis of stationarity; specifically, we adapt the co-integration tests of Shin (1994) and Leybourne and McCabe (1994), which are themselves variants of the KPSS test. We employ the statistic

$$S := s^{-2} T^{-2} \sum_{t=1}^T \left( \sum_{i=1}^t \hat{e}_i \right)^2, \tag{8}$$

where  $s^2 := (T - 3)^{-1} \sum_{t=1}^T \hat{e}_t^2$  and  $\hat{e}_t$  are the OLS residuals from the fitted regression

$$y_t = \hat{\alpha}_y + \hat{\beta}_x x_{t-1} + \hat{\beta}_{\Delta x} \Delta x_t + \hat{e}_t, \quad t = 1, \dots, T, \tag{9}$$

where, as in Shin (1994), the regressor  $\Delta x_t$  is included in (9) to account for the possibility of correlation between  $\epsilon_{xt}$  and  $\epsilon_{yt}$  ( $h_{31} \neq 0$ ). Abstracting from the role of the regressor  $\Delta x_t$ , when  $\beta_z \neq 0$ , the residuals  $\hat{e}_t$  incorporate a contribution of the unincluded  $z_{t-1}$  term in (1), hence the persistence in  $z_{t-1}$  is passed to  $\hat{e}_t$ , and the statistic  $S$  is a test of  $\beta_z = 0$  against  $\beta_z \neq 0$ , rejecting for large values of  $S$ . Specifically, assuming  $c_z = 0$ , we can rewrite (1) as

$$y_t = \alpha_y + \beta_x x_{t-1} + r_{t-1} + \epsilon_{yt}, \tag{10}$$

where  $r_t = r_{t-1} + u_t$ , initialized at  $r_0 = \beta_z \alpha_z$  (on setting  $s_{z,0} = 0$  with no loss of generality) with innovations  $u_t = \beta_z \epsilon_{zt}$ . Testing the null of  $\beta_z = 0$  against  $\beta_z = g_z T^{-1}$  in (1) is then seen to be precisely the same problem as testing the null of  $V(u_t) =: \sigma_{uu} = 0$  against  $\sigma_{uu} = g_z^2 T^{-2} \sigma_{zz}$  in the context of (10), with  $g_z = 0$  under both nulls. If we temporarily assume that  $x_t$  is strictly exogenous and  $\epsilon_{yt}$  and  $\epsilon_{zt}$  are independent IID normal random variates, then  $S$  is the locally best invariant (to  $\alpha_y$ ,  $\alpha_x$ ,  $\alpha_z$ ,  $\beta_x$ , and  $\sigma_{yy}$ ) test of the null  $\sigma_{uu} = 0$  against the local alternative  $\sigma_{uu} = g_z^2 T^{-2} \sigma_{zz}$  in (10). As such, the statistic  $S$  is relevant for our testing problem where we seek to distinguish between  $\beta_z = 0$  and  $\beta_z \neq 0$ . In our model we do not impose  $c_z = 0$  (nor the other temporary assumptions above), so in these more general circumstances we consider  $S$  to deliver a near locally best invariant test.

Notwithstanding the foregoing motivation, it is important to stress that a test based on  $S$  should properly be viewed as a misspecification test for the linear regression in (9). As such, a rejection by this test indicates that the fitted regression in (9) is not a valid PR. As with the failure of any misspecification test, this does not tell us why the regression has failed. We do know that  $S$  delivers a test which is (approximately) locally optimal in the direction of  $z_{t-1}$  being an unincluded variable (be it manifest or latent), but a rejection does not mean that  $x_{t-1}$  is not a valid predictor for  $y_t$ . Therefore, our proposed test is one for the invalidity of the putative PR, not of the putative predictor,  $x_{t-1}$ ; see again the discussion on this point in Section 2.

In Theorem 2 we now detail the limiting distribution of  $S$  under Assumption 1.

*Theorem 2.* For the DGP (1), (2), (3) and under Assumption 1,

$$S \xrightarrow{w} \int_0^1 \{F(r, c_x) + g_z G(r, c_x, c_z)\}^2 dr, \quad (11)$$

where

$$\begin{aligned} F(r, c_x) &:= \mathbb{B}_{\eta, y|x}(r) - \int_0^1 \bar{B}_{\eta c, x}(s) dB_{\eta, y|x}(s) \\ &\quad \times \left\{ \int_0^1 \bar{B}_{\eta c, x}(s)^2 \right\}^{-1} \int_0^r \bar{B}_{\eta c, x}(s), \\ G(r, c_x, c_z) &:= \left( \frac{\omega_{zz}}{\omega_{y|x}} \right)^{1/2} \left\{ \int_0^r \bar{B}_{\eta c, z}(s) \right. \\ &\quad \left. - \frac{\int_0^1 \bar{B}_{\eta c, x}(s) B_{\eta c, z}(s)}{\int_0^1 \bar{B}_{\eta c, x}^2(s)} \int_0^r \bar{B}_{\eta c, x}(s) \right\} \end{aligned}$$

with  $\omega_{y|x} := \omega_{yy} - \omega_{xy}^2/\omega_{xx}$ ,  $\mathbb{B}_{\eta, y|x}(r) := B_{\eta, y|x}(r) - rB_{\eta, y|x}(1)$ ,  $r \in [0, 1]$ , and  $B_{\eta, y|x} := \omega_{y|x}^{-1/2} \{M_{\eta y} - \omega_{xy} \omega_{xx}^{-1} M_{\eta x}\}$  a standardized heteroscedastic Brownian motion independent of  $B_1$ .

*Remark 7.* Notice that the limit in (11) does not depend on  $h_{31}$  owing to the invariance of the residuals  $\hat{e}_t$  to this parameter arising from the presence of the regressor  $\Delta x_t$  in (9). In the special case  $c_x = c_z$ , the limit is also invariant to  $h_{21}$  (see Remark 2). In fact, as  $M_{\eta z} = \omega_{xz} \omega_{xx}^{-1} M_{\eta x} + \omega_{z|x}^{1/2} B_{\eta 2}$  for  $\omega_{z|x} := \omega_{zz} - \omega_{xz}^2/\omega_{xx}$ , in this case the equality of the decay rate in the Ornstein–Uhlenbeck processes  $M_{\eta c, x}$  and  $M_{\eta c, z}$  ensures that  $B_{\eta c, z|x} := \omega_{z|x}^{-1/2} \{M_{\eta c, z} - \omega_{xz} \omega_{xx}^{-1} M_{\eta c, x}\}$  equals the Ornstein–Uhlenbeck process  $B_{\eta c, 2}$  so  $G(r, c_x, c_z)$  reduces to

$$\begin{aligned} G(r, c_x, c_x) &= \left( \frac{\omega_{z|x}}{\omega_{y|x}} \right)^{1/2} \left\{ \int_0^r \bar{B}_{\eta c, 2}(s) \right. \\ &\quad \left. - \frac{\int_0^1 \bar{B}_{\eta c, x}(s) B_{\eta c, 2}(s)}{\int_0^1 \bar{B}_{\eta c, x}^2(s)} \int_0^r \bar{B}_{\eta c, x}(s) \right\}. \end{aligned}$$

The term  $g_z G(r, c_x, c_z)$  in (11) is key in enabling the test  $S$  to potentially distinguish between  $H_u, H_x$  and  $H_z, H_{xz}$ . Clearly if  $\omega_{z|x}/\omega_{y|x} \simeq 0$ , then such a test has low power. This occurs when  $\epsilon_{xt}$  and  $\epsilon_{zt}$  are highly correlated (so  $\omega_{z|x} \simeq 0$ , corresponding to the part of  $z_{t-1}$  that is not shared and therefore not removed by the regressor  $x_{t-1}$ , on average over  $t$ ), or more generally, when

$\epsilon_{zt}$  corrected for  $\epsilon_{xt}$  varies little relatively to  $\epsilon_{yt}$  corrected for  $\epsilon_{xt}$ . For  $c_x \neq c_z$  the limit of  $S$  depends on  $h_{21}$  as  $G(r, c_x, c_x) - G(r, c_x, c_z)$  is proportional to  $h_{21} h_{11}^{-1}$ .

*Remark 8.* Under  $H_u, H_x$ , where  $g_z = 0$ , the limit distribution of  $S$  in (11) simplifies to  $\int_0^1 F(r, c_x)^2$  and depends only on  $c_x$  and any unconditional heteroscedasticity present in  $\epsilon_t$ .

*Remark 9.* We have assumed thus far that the  $\epsilon_{xt}$  are serially uncorrelated, with  $e_t$  being an m.d.s. More generally we may consider a linear process assumption for  $\epsilon_{xt}$  of the form  $\epsilon_{xt} = \sum_{i=0}^{\infty} \theta_i v_{x, t-i}$  where  $v_{x, t}$  is the first element of  $HD_t e_t$  with the standard summability and invertibility conditions  $\sum_{i=0}^{\infty} i|\theta_i| < \infty$  and  $\sum_{i=0}^{\infty} \theta_i z^i \neq 0$  for all  $|z| \leq 1$ , respectively, satisfied. Under homoscedasticity, this would include all stationary and invertible ARMA processes. Notice that  $\epsilon_{yt}$  remains uncorrelated with the increments of  $x_t$  at all lags (i.e.,  $x_t$  is weakly exogenous with respect to  $\epsilon_{yt}$ ) under this structure. Here, it may be shown that the limiting results given in Theorem 2 and in Theorems 3–5 continue to hold provided we replace (9) in the calculation of  $S$  with the augmented variant

$$\begin{aligned} y_t &= \hat{\alpha}_y + \hat{\beta}_x x_{t-1} + \hat{\beta}_{\Delta x} \Delta x_t + \sum_{i=1}^p \hat{\delta}_i \Delta x_{t-i} + \hat{e}_t, \\ t &= p+1, \dots, T, \end{aligned} \quad (12)$$

where  $p$  satisfies the standard rate condition that  $1/p + p^3/T \rightarrow 0$ , as  $T \rightarrow \infty$ , and it is assumed that  $T^{1/2} \sum_{i=p+1}^{\infty} |\delta_i| \rightarrow 0$ , where  $\{\delta_i\}_{i=1}^{\infty}$  are the coefficients of the  $AR(\infty)$  process obtained by inverting the  $MA(\infty)$  for  $\epsilon_{xt}$ . Similarly to BD, we would also need to restrict the amount of serial dependence allowed in the conditional variances via the assumption that  $\sup_{i, j \geq 1} \|\tau_{ij}\| < \infty$ , where  $\tau_{ij} := E(e_t e_t' \otimes e_{t-i} e_{t-i}')$ , with  $\otimes$  denoting the Kronecker product. Serial correlation of a similar form in  $\epsilon_{zt}$  will have no impact on our large-sample results under the null hypothesis,  $H_u, H_x$ , although an effect does arise under  $H_z, H_{xz}$ . As is standard in the PR literature, we maintain the assumption that  $\epsilon_{yt}$  is serially uncorrelated.

*Remark 10.* Extensions to the case where the putative PR contains multiple regressors and/or more general deterministic components can easily be handled in the context of our proposed PR invalidity test. Specifically, denoting the deterministic component as  $\tau' \mathbf{f}_t$ , where  $\mathbf{f}_t$  is as defined in Section 3.2 of BD, an obvious example being the linear trend case where  $\mathbf{f}_t := (1, t)'$ , and the vector of putative regressors as  $\mathbf{x}_{t-1}$ , then we would need to correspondingly construct  $S$  using the residuals from the regression of  $y_t$  on  $\mathbf{f}_t, \mathbf{x}_{t-1}$ , and  $\Delta \mathbf{x}_{t-1}$ . Doing so would alter the form of the limit distributions given in Theorem 2 and in the sequel, but would not alter the primary conclusion given in Corollary 1, that the fixed regressor wild bootstrap implementation of this test is asymptotically valid.

A consequence of the result in Theorem 2 is therefore that if we wish to base a test for PR invalidity on  $S$ , then we need to address the fact that under the null  $H_u, H_x$  the limit distribution of  $S$  is not pivotal. To account for the dependence of inference on any unconditional heteroscedasticity present, we employ a wild bootstrap procedure based on the residuals  $\hat{e}_t$ . However, we also need to account for the dependence of the limit distribution of  $S$  on  $c_x$ , and this we carry out by using the observed outcome on



$x := [x_0, \dots, x_T]'$  as a fixed regressor in the bootstrap procedure which we detail next.

### 4.2 A Fixed Regressor Wild Bootstrap Stationarity Test

A standard approach to obtaining bootstrap critical values for  $S$  would involve repeated generation of bootstrap samples for the original  $y_t$ , such that they mimic (in a statistical sense) the behavior of  $y_t$  with the null  $H_u, H_x$  imposed, together with repeated generation of bootstrap samples for the original  $x_t$ , to mimic the behavior of  $x_t$ . For each bootstrap sample, these would then be used to calculate a bootstrap analog of  $S$ , which should reflect the behavior of  $S$  under the null. Generation of bootstrap samples of  $y_t$  with suitable properties is quite straightforward, at least in large samples, using a standard wild bootstrap resampling scheme from the residuals  $\hat{e}_t$  from (9). However, finding bootstrap samples of  $x_t$  presents a significant problem since  $x_t = (1 - c_x T^{-1})x_{t-1} + \epsilon_{xt}$  (assuming  $\alpha_x = 0$  for simplicity) and so any corresponding recursion used to construct bootstrap samples for  $x_t$  from bootstrap samples of  $\epsilon_{xt}$  requires, for a size-controlled test, that  $c_x$  should be known or consistently estimated. Unfortunately, it is well-known that consistent estimation of  $c_x$  is not feasible. To avoid this problem, we circumvent estimation of  $c_x$  altogether and instead follow the approach taken in Hansen (2000), considering a bootstrap procedure which uses  $x$  as a fixed regressor, that is, the bootstrap statistic  $S^*$  is calculated from the *same* observed  $x_t$  as was used in the construction of  $S$  itself.

We now outline the steps involved in our proposed fixed regressor wild bootstrap.

*Algorithm 1 (Fixed Regressor Wild Bootstrap):*

- (i) Construct the wild bootstrap innovations  $y_t^* := \hat{e}_t w_t$ , where  $w_t, t = 1, \dots, T$ , is an IID  $N(0, 1)$  sequence independent of the data and  $\hat{e}_t$  are the residuals from either (9) or (12).
- (ii) Calculate the fixed regressor wild bootstrap analog of  $S$ ,

$$S^* := (s_y^*)^{-2} T^{-2} \sum_{t=1}^T \left( \sum_{i=1}^t \hat{\epsilon}_{yi}^* \right)^2,$$

where  $(s_y^*)^2 := (T - 2)^{-1} \sum_{t=1}^T (\hat{\epsilon}_{yt}^*)^2$  and  $\hat{\epsilon}_{yt}^*$  are OLS residuals from the fitted regression

$$y_t^* = \hat{\alpha}_y^* + \hat{\beta}_x^* x_{t-1} + \hat{\epsilon}_{yt}^*, \quad t = 1, \dots, T. \quad (13)$$

- (iii) Define the corresponding  $p$ -value as  $P_T^* := 1 - G_T^*(S)$  with  $G_T^*$  denoting the conditional (on the original data) cumulative distribution function (cdf) of  $S^*$ . In practice,  $G_T^*$  is unknown, but can be approximated in the usual way by numerical simulation.
- (iv) The wild bootstrap test of  $H_u, H_x$  at level  $\xi$  rejects in favor of  $H_z, H_{xz}$  if  $P_T^* \leq \xi$ .

*Remark 11.* The wild bootstrap scheme used to generate  $y_t^*$  is constructed so as to replicate the pattern of heteroscedasticity present in the original innovations; this follows because, conditionally on  $\hat{e}_t, y_t^*$  is independent over time with zero mean and variance  $\hat{e}_t^2$ .

*Remark 12.* By definition, the residuals  $\hat{e}_t$  from (9) are invariant to the value of  $\beta_x$  in (1), and so we can assume that  $\beta_x = 0$  with no loss of generality when generating the bootstrap  $y_t^*$  data. We also do not include  $\Delta x_t$  as an additional regressor (or lags thereof in the case considered in Remark 9) in (13) because the  $\hat{e}_t$  are asymptotically free of any effects arising from correlation between  $\epsilon_{xt}$  and  $\epsilon_{yt}$ , or from any weak dependence in  $\epsilon_{xt}$ .

*Remark 13.* Although  $\hat{e}_t$  depends on  $g_z$  under  $H_z, H_{xz}$ , we show in the next subsection that this does not translate into large-sample dependence of  $S^*$  on  $g_z$ .

### 4.3 Conditional Asymptotics and Bootstrap Validity

We show that the use of  $x_{t-1}$  as a fixed regressor in the construction of the bootstrap statistic  $S^*$  prevents  $S^*$  from converging weakly in probability to any nonrandom distribution, in contradistinction to most standard bootstrap applications we are aware of. Rather, under Assumption 1 and any of the hypotheses  $H_u, H_x, H_z$ , and  $H_{xz}$ , the distribution of  $S^*$ , given the data, converges weakly to the random distribution which obtains by conditioning the limit in (11) corresponding to  $g_z = 0$ , on the weak limit  $B_1$  of the process  $T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} e_{1t}$ ,  $r \in [0, 1]$ . This fact (along with some regularity conditions) makes it possible to conclude that the bootstrap  $p$ -value  $P_T^*$  is asymptotically uniform  $U[0, 1]$ -distributed under  $H_u, H_x$ , by using a general result on bootstrap validity from Cavaliere and Georgiev (2017, Theorem 2). From a pragmatic perspective, such a conclusion ensures that the bootstrap test is asymptotically size controlled under the conditions of Assumption 1 alone.

However, under Assumption 1 alone, the shortcoming remains that the meaning of the large-sample inference performed by our bootstrap test is unclear. Certainly, asymptotic bootstrap inference is not unconditional because  $S^*$  given the data does not converge to the unconditional limit distribution of  $S$ . On the other hand, bootstrap inference need not be asymptotically equivalent to conditional inference on  $x$  either. Indeed, it is well known that Theorem 2, where the limit distribution of  $S$  is established, cannot be taken to imply that  $S$  conditional on  $x$  converges weakly to the limit in (11) conditioned on  $B_1$  (the implication is falsified by, for example, Example 1 of LePage, Podgórski, and Ryznar 1997). Nevertheless, it is not unreasonable to expect that this result holds true under certain additional requirements, and we prove that this is in fact the case. We strengthen Assumption 1, so that under  $H_u, H_x$  the distribution of the statistic  $S$  conditional on  $x$  converges weakly to the same random distribution as  $S^*$  given the data, which allows us to establish that our bootstrap test in large samples has the meaning of a test conditional on  $x$ .

The results we present differ from those given by Hansen (2000) who considers a joint structural stability test on the constant and slope parameters in a general regression setting; our test of  $\beta_z = 0$  for the PR in (5) can be seen as the corresponding individual test for stability of just the intercept. Hansen argues that, under his Assumption 2, the fixed regressor (wild) bootstrap asymptotically implements unconditional inference (see Theorems 5 and 6, Hansen 2000) and that the convergence  $P_T^* \xrightarrow{w} U[0, 1]$  of bootstrap  $p$ -values under the null hypothesis follows from the equivalence of the unconditional limiting null

distribution of the original statistic and the limiting distribution of the bootstrap statistic given the data (see Corollaries 1 and 2, Hansen 2000). The results given in this section show that any such claim about unconditional inference is not correct, at least for the nonempty class of models satisfying both Hansen’s and our assumptions. Nonetheless the stated convergence of bootstrap  $p$ -values is correct, albeit for a different reason. A fuller treatment of this specific issue is given by Georgiev, Harvey, Leybourne, and Taylor (2018).

**Theorem 2** is based on the invariance principle given in (4). Conditional and bootstrap analogs of that theorem can be based on a *conditional* joint invariance principle for the original and the bootstrap data. To obtain this result, we will strengthen **Assumption 1** as follows:

*Assumption 2.* Let **Assumption 1** hold, together with the following conditions:

- (a)  $e_t$  is drawn from a doubly infinite strictly stationary and ergodic sequence  $\{e_t\}_{t=-\infty}^{\infty}$ , which is a martingale difference w.r.t. its own past.
- (b)  $\{[e_{2t}, e_{3t}]\}_{t=-\infty}^{\infty}$  is an m.d.s. also w.r.t.  $\mathcal{X} \vee \mathcal{F}_t$ , where  $\mathcal{X}$  and  $\mathcal{F}_t$  are the  $\sigma$ -algebras generated by  $\{e_{1t}\}_{t=-\infty}^{\infty}$  and  $\{[e_{2s}, e_{3s}]\}_{s=-\infty}^t$ , respectively, and  $\mathcal{X} \vee \mathcal{F}_t$  denotes the smallest  $\sigma$ -algebra containing both  $\mathcal{X}$  and  $\mathcal{F}_t$ .
- (c) The initial values  $s_{x,0}$  and  $s_{z,0}$  are measurable w.r.t.  $\mathcal{X}$  (in particular, they could be fixed constants).

*Remark 14.* Arguably, the most restrictive condition in **Assumption 2** is given in part (b). A first leading example where it is satisfied is that of a symmetric multivariate GARCH process with neither leverage nor asymmetric clustering. Specifically, let  $e_t = \Omega_t^{1/2} \varepsilon_t$ , where  $\Omega_t$  is measurable with respect to the past  $[\varepsilon_{1s}^2, \varepsilon_{2s}^2, \varepsilon_{3s}^2]'$ ,  $s \leq t - 1$ , and  $\{\varepsilon_t\}_{t=-\infty}^{\infty}$  is an iid sequence such that  $E(\varepsilon_{it} | \varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t}) = 0, i = 2, 3$ . If  $E\|e_t\| < \infty$ , then it could be seen that  $E(e_{it} | \mathcal{X} \vee \mathcal{F}_{t-1}) = 0, i = 2, 3$ . Another example is that of a multivariate stochastic volatility process  $e_t = H_t^{1/2} \varepsilon_t$  with  $\{H_t\}_{t=-\infty}^{\infty}$  independent of  $\{\varepsilon_t\}_{t=-\infty}^{\infty}$  and where  $\{\varepsilon_t\}_{t=-\infty}^{\infty}$  is an iid sequence with  $E(\varepsilon_{it} | \varepsilon_{1t}) = 0, i = 2, 3$  (which is certainly true if  $\varepsilon_t$  is multivariate standard Gaussian, as is usually assumed in the stochastic volatility framework). If  $E\|e_t\| < \infty$ , then again  $E(e_{it} | \mathcal{X} \vee \mathcal{F}_{t-1}) = 0, i = 2, 3$ . These two examples are also the leading examples given in the univariate context by Deo (2000), and in sec. 3 of Gonçalves and Kilian (2004). It would be interesting, although beyond the scope of our article, to investigate how **Assumption 2(b)** could be weakened to the case where  $\{e_t\}$  could be well approximated by a sequence satisfying **Assumption 2(b)**. For instance, following Rubshtein (1996), the conclusions of Theorem 5 in the supplementary appendix would remain valid if **Assumption 2(b)** was replaced by the condition that  $\sup_{t \geq 1} E\{E(\sum_{s=1}^t e_{is} | \mathcal{X})^2 < \infty, i = 2, 3$ .

In **Theorem 3**, we now establish three things: first, a conditional invariance principle that can be assembled from results and ideas disseminated throughout the probabilistic literature (see, in particular, Awad 1981; Rubshtein 1996), second, a bootstrap extension of that result, and third, associated convergence results for stochastic integrals. For simplicity, a one-dimensional bootstrap partial-sum process is considered; it is constructed from quantities  $\tilde{e}_{Tt}$  that we shall subsequently

specify to be the residuals  $\hat{e}_t$  from the regression in (9). Analogously to the definition of  $x$ , let  $y := [y_1, \dots, y_T]'$  and  $z := [z_0, \dots, z_T]'$ .

**Theorem 3.** Let  $\tilde{e}_{Tt}$  ( $t = 1, \dots, T$ ) be scalar measurable functions of  $x, y, z$  and such that  $\sum_{t=1}^{\lfloor Tr \rfloor} \tilde{e}_{Tt}^2 \xrightarrow{P} \int_0^r m^2(s) ds$  for  $r \in [0, 1]$ , where  $m$  is a square-integrable real function on  $[0, 1]$ . Introduce  $\tilde{\varepsilon}_{tb} := w_t \tilde{e}_{Tt}$  ( $t = 1, \dots, T$ ), and  $\tilde{B}_\eta(r) := \int_0^r m(s) d\tilde{B}_1(s)$ ,  $r \in [0, 1]$ , where  $\tilde{B}_1$  is a standard Brownian motion independent of  $B$ . Under **Assumption 2**, the following converge jointly as  $T \rightarrow \infty$ :

$$\left( T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t, T^{-1} \sum_{t=1}^T \sum_{s=1}^{t-1} \varepsilon_{xs} [\varepsilon_{yt}, \varepsilon_{zt}] \right) \Big| x$$

$$\xrightarrow{w} \left( M_\eta(r), \int_0^1 M_{\eta x}(s) d[M_{\eta y}(s), M_{\eta z}(s)] \right) \Big| B_1,$$

$r \in [0, 1]$ , in the sense of weak convergence of random measures on  $\mathcal{D}^3 \times \mathbb{R}^2$ , and

$$\left( T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} [e_{1t}, \tilde{\varepsilon}_{tb}], T^{-1} \sum_{t=1}^T \sum_{s=1}^{t-1} \varepsilon_{xs} \tilde{\varepsilon}_{tb} \right) \Big| x, y, z$$

$$\xrightarrow{w} \left( B_1(r), \tilde{B}_\eta(r), \int_0^1 M_{\eta x}(s) d\tilde{B}_\eta(s) \right) \Big| B_1,$$

$r \in [0, 1]$ , in the sense of weak convergence of random measures on  $\mathcal{D}^2 \times \mathbb{R}$ .

*Remark 15.* Let  $E_x(\cdot) := E(\cdot | x)$  and  $E^*(\cdot) := E(\cdot | x, y, z)$ . The convergence concept used in **Theorem 3** is defined as follows. Let  $\zeta, \zeta_T$  and  $\xi, \xi_T$  ( $T \in \mathbb{N}$ ) be random elements of the metric spaces  $\mathcal{S}$  and  $\mathcal{T}$ , respectively, such that  $\zeta, \xi$  and  $B_1$  are defined on the same probability space, and similarly for  $\zeta_T, \xi_T$  and  $x, y, z$ . We say that  $\zeta_T | x \xrightarrow{w} \zeta | B_1$  and  $\xi_T | x, y, z \xrightarrow{w} \xi | B_1$  jointly in the sense of weak convergence of random measures on  $\mathcal{S}$  and  $\mathcal{T}$  if for all bounded continuous functions  $f : \mathcal{S} \rightarrow \mathbb{R}$  and  $g : \mathcal{T} \rightarrow \mathbb{R}$  it holds that

$$[E_x(f(\zeta_T)), E^*(g(\xi_T))] \Big| \xrightarrow{w} [E(f(\zeta) | B_1), E(g(\xi) | B_1)]'$$

as  $T \rightarrow \infty$ , in the sense of standard weak convergence of random vectors in  $\mathbb{R}^2$ .

We are already in a position to establish in **Theorem 4** the large-sample behavior of  $S$  conditional on  $x$ , and of  $S^*$ , its bootstrap analog from **Algorithm 1**, conditional on the data. These two limiting distributions will be seen to coincide under the null hypothesis.

**Theorem 4.** Under DGP (1)–(3) and **Assumption 2**, the following converge jointly as  $T \rightarrow \infty$ , in the sense of weak convergence of random measures on  $\mathbb{R}$ :

$$S | x \xrightarrow{w} \int_0^1 \{F(r, c_x) + g_z G(r, c_x, c_z)\}^2 dr \Big| B_1 \quad (14)$$

$$S^* | x, y, z \xrightarrow{w} \int_0^1 F(r, c_x)^2 dr \Big| B_1, \quad (15)$$

where the processes  $F$  and  $G$  are as defined in **Theorem 2**.

*Remark 16.* A comparison of (14) and (15) shows that the bootstrap statistic  $S^*$ , conditional on the data, and the original statistic  $S$ , conditional on  $x$ , converge jointly to the same random distribution when  $g_z = 0$ , that is, under the null hypothesis,  $H_u, H_x$ . An implication of this is that the bootstrap approximation is consistent in the sense that

$$\sup_{u \in \mathbb{R}} |P_x(S \leq u) - P^*(S^* \leq u)| \xrightarrow{P} 0, \tag{16}$$

given that the random cdf of  $\int_0^1 F(r, c_x)^2 dr | B_1$  is sample-path continuous. Here  $P_x$  and  $P^*$  denote probability conditional on  $x$  and on all the data, respectively. Thus, the distribution of the “fixed-regressor bootstrap” statistic  $S^*$  conditional on the data consistently estimates the large-sample distribution of the original statistic  $S$  conditional on the “fixed regressor”  $x$ . This result differs from the usual formulation of bootstrap validity, where two cdfs with a common nonrandom limit are compared; here, in contrast,  $P_x(S \leq u) \xrightarrow{w} P(\int_0^1 F(r, c_x)^2 dr \leq u | B_1)$ ,  $u \in \mathbb{R}$ , with a nondegenerate random limit.

In Corollary 1, we formulate the conclusion of asymptotic validity of the bootstrap test based on  $S$  and  $S^*$  in terms of the bootstrap  $p$ -values.

*Corollary 1.* Let  $P_T^* := P^*(S^* > S)$ . Under  $H_u, H_x$  and Assumption 2,  $P_T^* | x \xrightarrow{w} U[0, 1]$  and  $P_T^* \xrightarrow{w} U[0, 1]$ .

An implication of Corollary 1 is that comparison of the statistic  $S$  with a  $\xi$  level bootstrap critical value (approximated by the upper tail  $\xi$  percentile from the order statistic formed from  $B$  independent simulated bootstrap  $S^*$  statistics, which we will denote by  $cv_{\xi, B}$ ) results in a bootstrap test with correct asymptotic size ( $\xi$ ) under  $H_u, H_x$ , conditionally on  $x$  and unconditionally. In what follows we denote by  $S_B$  the fixed regressor wild bootstrap procedure outlined in Algorithm 1, whereby  $S$  is compared to the critical value  $cv_{\xi, B}$ . The asymptotic local power of  $S_B$  under  $H_z, H_{xz}$  depends on the parameter  $g_z$ .

*Remark 17.* For the bootstrap statistic,  $S^*$ , the same limiting distribution is obtained in (15) under the alternative hypothesis,  $H_z, H_{xz}$ , as under the null hypothesis. In contrast, in the case of  $S$ , a stochastic offset, arising from the term  $g_z G(r, c_x, c_z)$ , is seen in the limiting distributions (in (14) conditionally on  $x$ , and in (11) unconditionally). Although, for a given alternative, the asymptotic local power is different for the bootstrap test based on  $S^*$  and an (infeasible) test based on the unconditional limit of  $S$  and knowledge of the parameter  $c_x$  (the former power is a random variable depending on  $B_1$  and the latter power is a number), we comment in Remark 18 on some qualitative similarities.

*Remark 18.* The limiting functional for  $S$  in (11) and (14) is dominated in probability (both unconditionally and conditionally on  $B_1$ ) by  $g_z^2 \int_0^1 G(r, c_x, c_z)^2 dr$  for large  $g_z$  and, as a result, asymptotic local power approaches 1 as  $g_z$  diverges. Nonetheless, asymptotic local power is not monotone in  $|g_z|$ . For example, in the case  $c_x = c_z$ , the null component  $F(r, c_x)$  in (11) and (14) involves a term in  $h_{32} B_{\eta 2}(r)$ , while the alternative component  $g_z G(r, c_x, c_z)$  involves a term in  $g_z \int_0^r \bar{B}_{\eta c, 2}$  (see Remark 7). Because  $B_{\eta 2}(r)$  and  $\int_0^r \bar{B}_{\eta c, 2}$  are positively correlated, it can be shown that  $E\{\int_0^1 F(r, c_x) G(r, c_x, c_z) dr\} \neq 0$  for  $h_{32} \neq 0$ , and similarly for the conditional expectation given  $B_1$ , a.s. As a

result, when  $h_{32} \neq 0$ , there exist values of  $g_z$  (dependent on  $B_1$  in the conditional case) which render the expectations of the limits in (11) and (14) (respectively, unconditional and conditional on  $B_1$ ), smaller than their expectations under the null hypothesis. For such  $g_z$  the limit distribution under the alternative does not first-order stochastically dominate the limit distribution under the null, translating into power being less than size for some size levels.

#### 4.4 Asymptotic Local Power of Stationarity Tests Under $H_z$

We now consider the asymptotic local power of  $S$  and  $S_B$ , the latter on average over  $B_1$ . We use the same set of homoscedastic simulation models as for the size of  $t_u$  and  $Q$  in Figure 1, so we overlay this information on them. For the asymptotic power of  $S$  under  $H_z$ , we use the limit expression (11), having first obtained 0.10-level critical values from simulating (11) under  $g_z = 0$ . Since these critical values depend on knowledge of  $c_x$ ,  $S$  here is an infeasible test against which to benchmark the power of  $S_B$ . The asymptotic power of  $S_B$  is also based on the limit distribution of  $S$  under  $H_z$  but compared against a simulated limit bootstrap critical value  $cv_{\xi, B}$  with  $\xi = 0.10$ . For each replication, this critical value is obtained by simulating the limit (15) using  $B = 2000$  replications, conditioning on the simulated  $B_1$  for that Monte Carlo replication.

When  $c = 0$ , we see the power of  $S$  rising rapidly with departures from  $g_z = 0$ . For  $g_z = 50$ , its power is very close to 1. Turning attention to  $S_B$ , it has a very similar power profile to that of  $S$ ; indeed, its power marginally exceeds that of  $S$ . It is of course anticipated from Remark 17 that  $S_B$  does not have the same asymptotic local power function as  $S$ , but the fact that its power exceeds that of  $S$  is a welcome finding as  $S_B$ , unlike  $S$ , is a feasible procedure. When  $c = 10$  the powers of  $S$  and  $S_B$  are near identical, but at a lower level than when  $c = 0$ . There is also a nonmonotonicity in the power profiles of  $S$  and  $S_B$ , anticipated from Remark 18, for  $\sigma_{zy} = -0.70$  when  $g_z$  is small, with power dipping below size. However, for large enough  $g_z$ , this anomaly disappears. (We note that  $S$  is not LBI when we allow correlation between  $\epsilon_{yt}$  and  $\epsilon_{zt}$  so this anomalous behavior is perhaps not entirely surprising.)

The important comparison here is between the power of  $S_B$  (restricting attention to the feasible procedure) and the size of  $t_u$  and  $Q$  (as their size profiles are similar we only refer to  $t_u$ ). When  $c = 0$ , the power of  $S_B$  exceeds the size of  $t_u$ , hence the invalidity of the PR is detected with greater frequency than  $t_u$  spuriously rejects in favor of predictability of  $y_t$  by  $x_{t-1}$ . This demonstrates the capability of  $S_B$  to detect PR invalidity in cases where the important size problems associated with  $t_u$  exist. That the power of  $S_B$  exceeds the size of  $t_u$  under  $H_z$  is possibly to be expected, because  $S$  is designed to detect departures from the null of  $g_z = 0$  whereas such departures simply represent model misspecification in the context of the PR test  $t_u$ . With  $c = 10$ , we again see that the power of  $S_B$  generally out-strips the sizes of  $t_u$ , with the size/power differences appearing even more marked than for  $c = 0$ . Again, the only exception to this is for  $\sigma_{zy} = -0.7$  when  $g_z$  is small.



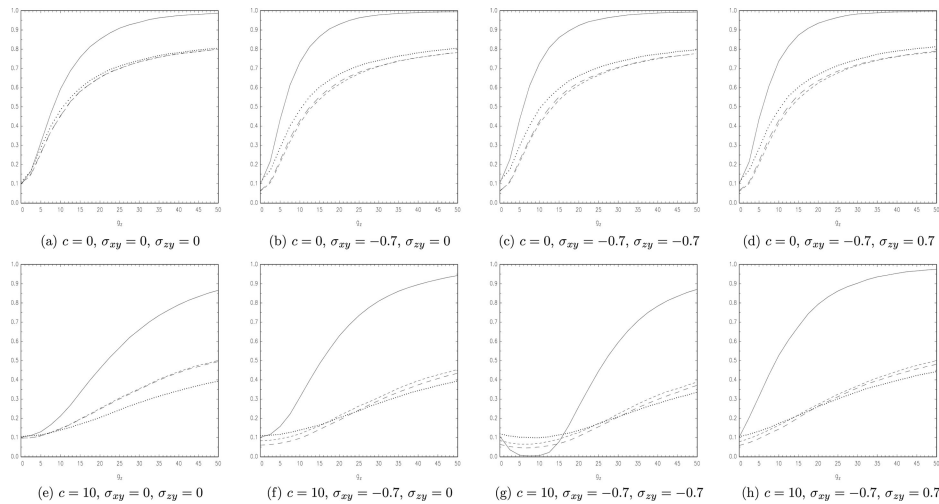


Figure 2. Finite sample rejection frequencies of  $S_B$  (power) and  $t_u$ ,  $Q$ ,  $IV_{\text{comb}}$  (size):  $T = 200$ ,  $g_x = 0$ ,  $c_x = c_z = c$ ;  $S_B$ : —,  $t_u$ : - - -,  $Q$ : - · -,  $IV_{\text{comb}}$ : · · ·

The supplementary appendix to this article contains asymptotic power simulation results for some additional parameter configurations (for which many possibilities exist). We consider the current setup with  $c = 5$  and  $c = 20$  and we find that the power of  $S_B$  with  $c = 20$  is lower than for  $c = 10$  due to a less persistent  $z_{t-1}$  lessening the impact of model misspecification. Other simulations where we allow  $c_z$  to be different to  $c_x$  confirm that the main driver of power for  $S_B$  is  $c_z$  and not  $c_x$ , as would be expected. We also consider  $\sigma_{xz} \neq 0$  (with  $c_z$  and  $c_x$  equal or different; note that we reduce the magnitudes of  $\sigma_{xy}$  and  $\sigma_{zy}$  in some cases to ensure  $\Omega$  remains positive definite). Here the interplay between  $S_B$  and  $t_u$  ( $Q$ ) becomes rather more complex. For example, with  $c_z = c_x$ , setting  $\sigma_{xz} = \pm 0.5$  causes the power of  $S_B$  to suffer while the frequency with which  $t_u$  rejects increases, while for  $c_z \neq c_x$ , only small changes are observed for  $\sigma_{xz} \neq 0$  compared to  $\sigma_{xz} = 0$ .

## 5. FINITE SAMPLE SIZE AND POWER UNDER $H_z$

We now evaluate the finite sample size properties of the PR tests and the size and power of  $S_B$ . For the PR tests, we consider the feasible versions of  $t_u$  and  $Q$ , proposed by CES and CY, respectively, both of which rely on Bonferroni bounds to control size. (We are grateful to Campbell and Yogo for making their Gauss code available for these two procedures.) We also consider the IV-based test of BD that combines fractional and sine function instruments, denoted  $IV_{\text{comb}}$ , comparing this with its asymptotic  $\chi^2(1)$  critical value. For  $S_B$  we use  $B = 499$  replications.

To begin, we continue to abstract from heteroscedasticity and consider finite sample DGPs for the same settings as used in the main asymptotic simulations. Specifically, we simulate the DGP (1)–(3) for  $T = 200$  with  $\alpha_y = \alpha_x = \alpha_z = 0$ ,  $g_x = 0$ ,  $s_{x,0} = s_{z,0} = 0$ ,  $d_{it} = 1$  ( $i = 1, 2, 3$ ), and  $e_t \sim \text{IID } N(0, I_3)$ . Figure 2 reports the finite sample analogs of Figure 1, that is, rejection frequencies of nominal 0.10-level (two-sided for  $t_u$ ,  $Q$ , and  $IV_{\text{comb}}$ ) tests under  $H_z$ . Simulations are again conducted

using 10,000 Monte Carlo replications. On comparing Figure 2 with its large-sample counterpart Figures 1, it is clear that our asymptotic simulations provide a close approximation to the finite sample rejection frequencies of  $t_u$ ,  $Q$ , and  $S_B$ , particularly in terms of the relative behavior of the tests, albeit in absolute terms the finite sample rejection frequencies tend to be slightly lower than their asymptotic counterparts. For  $t_u$  and  $Q$ , this is partly due to the feasible tests not having the same large-sample properties as the infeasible tests. The general observations made on the basis of the asymptotic simulations apply equally here; finite sample size of the PR tests increases with  $g_z$ , giving rise to an increasing likelihood of concluding spurious predictive ability. As anticipated in the discussion of Section 3.1, a similar pattern of rejections is found for  $IV_{\text{comb}}$ ; its sizes are close to those of  $t_u$  and  $Q$ . As regards  $S_B$ , its finite sample power increases with  $g_z$ , with the invalidity of the PR generally being detected with greater frequency than the PR tests' spurious rejections. Hence, the ability of  $S_B$  to detect PR invalidity in cases where well-known PR tests suffer problematic over-size is displayed in finite samples also.

Finally, we examine the impact of unconditional heteroscedasticity in the DGP on the size of  $S_B$  and  $IV_{\text{comb}}$  when the error processes are subject to a single break in volatility. (We do not consider  $t_u$  and  $Q$  here since these procedures are not robust to heteroscedastic errors.) Specifically, we again simulate the DGP (1)–(3) for  $T = 200$  with  $g_x = g_z = 0$ ,  $e_t \sim \text{IID } N(0, I_3)$ , but setting  $d_{it} = \mathbb{I}(t \leq \lfloor \tau T \rfloor) + \sigma_i \mathbb{I}(t > \lfloor \tau T \rfloor)$  for  $i = 1, 3$ . We set  $\tau = \{0.3, 0.7\}$  thereby allowing for two (common) volatility break timings, and  $\sigma_i = \{1, 4, \frac{1}{4}\}$  allowing for both upward and downward volatility shifts (these magnitudes being substantial for illustrative purposes). We consider  $c_x = \{0, 5, 10\}$  and for simplification abstract from time-varying correlation between  $\epsilon_{xt}$  and  $\epsilon_{yt}$  by setting  $h_{21} = h_{31} = h_{32} = 0$ . Table 1 reports the results for nominal 0.10-level tests (two-sided for  $IV_{\text{comb}}$ ). It is clear that the size of  $S_B$  is very well controlled across all the patterns of time-varying volatility of  $\epsilon_{xt}$  and  $\epsilon_{yt}$ . The wild bootstrap aspect of the bootstrap methods that we propose therefore works well in achieving size close to the nominal level even for



Table 1. Finite sample size of  $S_B$  and  $IV_{comb}$  under volatility shifts:  $T = 200, g_x = g_z = 0, d_{it} = 1(t \leq \lfloor \tau T \rfloor) + \sigma_i 1(t > \lfloor \tau T \rfloor), i = 1, 3$

		$c_x = 0$				$c_x = 5$				$c_x = 10$			
		$\tau = 0.3$		$\tau = 0.7$		$\tau = 0.3$		$\tau = 0.7$		$\tau = 0.3$		$\tau = 0.7$	
$\sigma_1$	$\sigma_3$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$
1	1	0.098	0.110	0.098	0.110	0.103	0.104	0.103	0.104	0.102	0.105	0.102	0.105
	4	0.101	0.109	0.101	0.112	0.106	0.107	0.105	0.111	0.105	0.108	0.107	0.110
	$\frac{1}{4}$	0.102	0.112	0.098	0.104	0.104	0.105	0.099	0.105	0.104	0.106	0.102	0.105
4	1	0.100	0.109	0.102	0.113	0.103	0.107	0.104	0.112	0.104	0.108	0.104	0.113
	4	0.099	0.109	0.102	0.117	0.107	0.110	0.107	0.119	0.106	0.114	0.109	0.123
	$\frac{1}{4}$	0.101	0.107	0.099	0.099	0.104	0.102	0.102	0.100	0.106	0.102	0.102	0.103
$\frac{1}{4}$	1	0.102	0.114	0.099	0.111	0.102	0.108	0.105	0.107	0.104	0.109	0.110	0.106
	4	0.103	0.105	0.103	0.108	0.102	0.100	0.108	0.106	0.104	0.100	0.108	0.105
	$\frac{1}{4}$	0.103	0.117	0.098	0.108	0.105	0.112	0.101	0.108	0.106	0.113	0.101	0.110

the large volatility changes that we consider. (We also simulated the finite sample size of  $S_B$  under a variety of conditionally heteroscedastic specifications, including multivariate GARCH and EGARCH, the latter an example of an asymmetric GARCH process. The size of  $S_B$  was found to be well controlled, with only minor deviations from the nominal level.) The  $IV_{comb}$  test also displays a good degree of robustness to heteroscedasticity, although size can be a little inflated for some settings.

The supplementary appendix also contains results for the same settings as above but with  $g_z = 25$  and  $g_z = 50$ , that is, power for  $S_B$  and size for  $IV_{comb}$ , with  $c_z = c_x$  and additionally allowing for a volatility break in  $\epsilon_{zt}$  via  $d_{zt} = \mathbb{I}(t \leq \lfloor \tau T \rfloor) + \sigma_2 \mathbb{I}(t > \lfloor \tau T \rfloor)$ . It is clear that the presence of (unconditional) heteroscedasticity can have a substantial influence on the level of power attainable. Other things equal, a volatility increase in  $\epsilon_{zt}$  (an increase in  $\sigma_2$ ) leads to higher  $S_B$  power, with a volatility decrease in  $\epsilon_{zt}$  having the opposite effect, while volatility changes in  $\epsilon_{yt}$  have the reverse effect, with an increase (decrease) in  $\sigma_3$  resulting in lower (higher) power for  $S_B$ . Volatility changes in  $\epsilon_{xt}$  (changes in  $\sigma_1$ ) appear to have relatively little effect. A similar pattern of rejection frequencies is also observed for the sizes of the  $IV_{comb}$  test under heteroscedasticity. In the same cases where  $S_B$  power is increased (decreased), so the over-size of  $IV_{comb}$  increases (decreases). It appears, therefore, that  $S_B$  has attractive size and power properties in finite samples as well as in the limit, and it is encouraging to see that for the most part these carry over to situations where the errors are unconditionally heteroscedastic.

## 6. AN EMPIRICAL APPLICATION TO U.S. EQUITY DATA

To illustrate how our proposed procedure may be used in practice, we reconsider the results from the empirical analysis investigating the predictability of excess returns using the U.S. equity data reported in CY. CY consider four different series of stock returns, dividend-price ratio, and earnings-price ratio. The first is annual S&P 500 index data over the period 1871–2002. The other three series are annual, quarterly, and monthly NYSE/AMEX

value-weighted index data (1926–2002). Full data descriptions are provided in CY. The data can be obtained from <https://sites.google.com/site/motohiroyogo/home/research/>

CY analyze the time series behavior of these data and test for predictability in excess returns (relative to an appropriate risk free rate), using as putative predictors for a variety of sample windows: the dividend-price ratio, denoted  $d - p$ ; the earnings-price ratio, denoted  $e - p$ ; the three-month T-bill rate, denoted  $r_3$ , and a measure of the long-short yield spread, denoted  $y - r_1$ . Details on the construction of these variables can be found in CY; as is conventional, excess returns and the predictor variables appear in logs. CY argue that all of these possible predictors display high persistence with, in most cases, the 95% confidence interval for the largest autoregressive root containing unity. A priori then, bivariate tests of predictability would seem to be at potential risk from the spurious predictability problem.

Table 2 reports the application of a variety of statistics to the same sets of bivariate PRs as in Table 5 of CY. Here  $S$  is our PR invalidity statistic; KPSS is the KPSS for stationarity of the predictor appearing in that regression;  $IV_{comb}$  is the PR test of BD. The  $S$  statistic is implemented using BIC selection for the order of  $p$  in the fitted regression (12), starting from  $p_{max} = 12$ , with an appropriate degrees of freedom adjustment made for  $s_y^2$ . (We have simulated this means of selection of  $p$  across a number of different stationary ARMA DGPs for  $\epsilon_{xt}$  and it appears to control the size of  $S_B$  well.) For the KPSS statistic the long run variance estimate is based on the QS kernel with automatic bandwidth selection. For each test, a  $p$ -value is given. For  $S$  this relates to our fixed regressor wild bootstrap test,  $S_B$  using  $B = 9999$  replications; for KPSS it is based on the wild bootstrap method of Cavaliere and Taylor (2005), again using  $B = 9999$ ; for  $IV_{comb}$  it relates to a  $\chi^2(1)$  distribution. Finally, under  $Q$ , an entry of \* (NS) denotes that CY's  $Q$  test rejects (does not reject) the null of no predictability at the 0.10 level.

Notice first that the  $p$ -values for KPSS are relatively close to zero for most of the predictors. The KPSS test is known to reject the null of stationarity with high probability when a series displays local-to-unit root behavior (increasingly as the local-to-unity parameter approaches zero), so the  $p$ -value can be viewed as an indicator of the strength of persistence in a series (higher persistence associated with a lower  $p$ -value).

Table 2. Application to U.S. Equity Indices

Series	Obs.	Predictor	$S$	$p$ -Val.	KPSS	$p$ -Val.	$IV_{\text{comb}}$	$p$ -Val.	$Q$
Panel A: S&P 1880–2002, CRSP 1926–2002									
S&P 500	123	$d - p$	0.358	0.057	0.669	0.043	0.187	0.426	NS
		$e - p$	1.111	0.000	0.449	0.087	1.087	0.139	*
Annual	77	$d - p$	0.081	0.658	0.572	0.077	1.383	0.083	*
		$e - p$	0.522	0.008	0.465	0.116	0.988	0.162	*
Quarterly	305	$d - p$	0.531	0.017	1.201	0.007	0.474	0.319	NS
		$e - p$	1.302	0.000	0.889	0.026	0.624	0.267	*
Monthly	913	$d - p$	1.449	0.000	2.588	0.000	-0.423	0.337	NS
		$e - p$	1.522	0.000	1.938	0.001	-0.139	0.445	*
Panel B: S&P 1880–1994, CRSP 1926–1994									
S&P 500	115	$d - p$	0.346	0.081	0.495	0.028	0.388	0.350	NS
		$e - p$	1.207	0.000	0.251	0.146	1.600	0.054	*
Annual	69	$d - p$	0.100	0.611	0.390	0.062	1.593	0.055	*
		$e - p$	0.803	0.002	0.272	0.222	1.206	0.114	*
Quarterly	273	$d - p$	0.894	0.001	0.753	0.009	0.451	0.327	NS
		$e - p$	2.028	0.000	0.420	0.114	0.711	0.239	*
Monthly	817	$d - p$	1.626	0.000	1.473	0.000	-0.598	0.276	NS
		$e - p$	2.434	0.000	0.839	0.021	-0.164	0.435	*
Panel C: CRSP 1952–2002									
Annual	51	$d - p$	0.368	0.051	0.351	0.210	1.286	0.099	NS
		$e - p$	0.058	0.675	0.244	0.270	0.979	0.163	NS
Quarterly	204	$r_3$	0.071	0.726	0.269	0.151	-1.391	0.082	NS
		$y - r_1$	0.085	0.657	0.626	0.014	0.472	0.381	NS
		$d - p$	0.518	0.017	0.645	0.062	1.128	0.129	NS
		$e - p$	1.511	0.000	0.550	0.064	0.764	0.223	NS
Monthly	612	$r_3$	0.071	0.659	0.585	0.017	-2.661	0.004	*
		$y - r_1$	0.235	0.146	0.855	0.003	0.946	0.172	*
		$d - p$	0.345	0.073	1.449	0.004	0.550	0.290	NS
		$e - p$	1.729	0.000	1.264	0.004	0.363	0.358	NS
Monthly	612	$r_3$	0.091	0.535	1.296	0.000	-3.439	0.000	*
		$y - r_1$	0.422	0.028	1.373	0.000	1.856	0.032	*

NOTES: Returns are for the annual S&P 500 index and the annual, quarterly, and monthly CRSP value-weighted index. The predictor variables are the log dividend-price ratio  $d - p$ , the log earnings-price ratio  $e - p$ , the three-month T-bill rate  $r_3$ , and the long-short yield spread  $y - r_1$ . In the column headed  $Q$ , \* (NS) indicates those cases where the  $Q$  test of Campbell and Yogo (2006) rejects (does not reject) the null hypothesis of no predictability at the 10% level. The columns headed  $p$ -val. indicate the  $p$ -values of the tests in the preceding column calculated as detailed in the main text.

We conclude that, in accordance with the findings of CY and BD, these possible predictors all display (to differing degrees) strongly persistent behavior. The least persistent appears to be the annual log earnings-price ratio,  $e - p$ , regardless of which sample window is considered. Interestingly, while CY suggest that  $r_3$  and  $y - r_1$  are the least persistent variables, we find small  $p$ -values for these series in almost every case, suggesting they are strongly persistent.

For both the full sample results in Panel A and the sub-sample considered in Panel B, the  $Q$  test delivers rejections at the 0.10 level in the case of  $e - p$ , for all four of the data series considered. The  $Q$  test also rejects at the 0.10 level for  $d - p$ , but only for annual data. The  $IV_{\text{comb}}$  test also generally rejects with annual data. These results, when taken at face value, signal significant predictability of excess returns by  $e - p$  in particular, but also by  $d - p$  with annual data. However, in the case of  $e - p$  any such conclusions of predictability are immediately thrown into serious question once we observe that  $S_B$  also rejects very strongly in all these cases, suggesting that such a PR model

is potentially spurious, or at the very least, under-specified by some unincluded persistent process. Interestingly, in the annual data the  $S_B$  test for  $d - p$  is highly insignificant in both Panels A and B suggesting no evidence that the significant outcome of the  $Q$  test is spurious here. So although the evidence from the  $Q$  tests alone suggests that  $e - p$  has predictive power for excess returns with a less consistent body of evidence of predictability from  $d - p$ , a consideration of the  $Q$  tests in tandem with  $S_B$  suggests that the stronger evidence for genuine predictability may well lie with  $d - p$ ; indeed the results are not inconsistent with  $d - p$  being an omitted manifest persistent predictor when testing for predictability from  $e - p$ .

Turning to the results in Panel C, the  $Q$  test is seen to be significant at the 0.10 level only for  $r_3$  and  $y - r_1$  for quarterly and monthly, but not annual, data. Among these cases, only  $y - r_1$  for monthly data is flagged up as potentially spurious by  $S_B$ . Consequently, with this exception, the rejections delivered by  $Q$  in Panel C do not appear problematic when judged by our PR validity test. For the  $IV_{\text{comb}}$  test in Panel C, significant

predictability at the 0.10 level is again (as with  $Q$ ) signaled for monthly  $r_3$  and monthly  $y - r_1$ , but also signaled for annual  $d - p$  and both annual and quarterly  $r_3$ . The results for  $S_B$  again suggest that most of these rejections do not appear to be obviously problematic, although  $S_B$  does reject at roughly the 0.05 level for annual  $d - p$ .

## 7. CONCLUSIONS

In this article, we have examined the issue of spurious predictability that can potentially arise with recently proposed tests for predictability. We have shown that the outcomes from these tests have considerable potential to spuriously signal that a putative predictor is a genuine predictor whenever unincorporated persistent (manifest and/or latent) variables are present in the underlying data generation process. To guard against this possibility, we have proposed a diagnostic test for such PR invalidity based on a well-known stationarity testing approach. To again allow for an unknown degree of persistence in the putative (and latent) predictors, and to allow for both conditional and unconditional heteroscedasticity in the data, a fixed regressor wild bootstrap test procedure was proposed and its asymptotic validity established. Doing so required us to establish some novel asymptotic results pertaining to the use of the fixed regressor bootstrap with nonstationary regressors, which are likely to have important applications beyond the present context. Monte Carlo simulations were reported which suggested that our proposed methods work well in practice. A reconsideration of the empirical study of the predictability of U.S. stock returns reported in CY highlighted the potential value of our procedure in practice.

We have proposed what we believe to be the first serious diagnostic testing exercise in the context of fitted PRs, suggesting within-sample misspecification tests directed to have power to detect the presence of persistent variables in the underlying DGP but not included in the PR. We hope that this article encourages further research in this area, developing additional within- and out-of-sample diagnostic procedures for PRs.

## SUPPLEMENTARY MATERIALS

This supplement contains the additional Monte Carlo simulation results described in sections 4.4 and 5, together with mathematical proofs for the large sample results given in sections 3 and 4 of the article.

## ACKNOWLEDGMENTS

The authors are grateful to the editor, Todd Clark, an anonymous co-editor, and three anonymous referees for their helpful and constructive comments. The authors particularly thank one of the referees for suggesting the examples relating to latent predictors discussed in Section 2.

Taylor gratefully acknowledges financial support provided by the Economic and Social Research Council of the United Kingdom under research grant ES/R00496X/1.

## ORCID

A. M. Robert Taylor  <http://orcid.org/0000-0003-0567-0276>

## REFERENCES

- Awad, A. (1981), "Conditional Central Limit Theorems for Martingales and Reversed Martingales," *Sankhya*, 43 A, 100–106. [536]
- Bansal, R., and Yaron, A. (2004), "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *The Journal of Finance*, 59, 1481–1509. [530]
- Boswijk, P., Cavaliere, G., Rahbek, A., and Taylor, A. M. R. (2016), "Inference on Co-Integration Parameters in Heteroskedastic Vector Autoregressions," *Journal of Econometrics*, 192, 64–85. [531]
- Breitung, J., and Demetrescu, M. (2015), "Instrumental Variable and Variable Addition Based Inference in Predictive Regressions," *Journal of Econometrics*, 187, 358–375. [529]
- Cai, Z., Wang, Y., and Wang, Y. (2015), "Testing Instability in a Predictive Regression Model With Nonstationary Regressors," *Econometric Theory*, 31, 953–980. [530,532]
- Campbell, J. Y. and Yogo, M. (2006), "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, 81, 27–60. [528]
- Cavaliere, G., and Georgiev, I. (2017), "Bootstrap Inference Under Random Distributional Limits," Research Institute for Econometrics Discussion Paper 1-17, Bar-Ilan University. [535]
- Cavaliere, G., Rahbek, A., and Taylor, A. M. R. (2010), "Testing for Co-Integration in Vector Autoregressions With Non-Stationary Volatility," *Journal of Econometrics*, 158, 7–24. [530]
- Cavaliere, G., and Taylor, A. M. R. (2005), "Stationarity Tests Under Time-Varying Second Moments," *Econometric Theory*, 21, 1112–1129. [539]
- Cavanagh, C. L., Elliott, G., and Stock, J. H. (1995), "Inference in Models With Nearly Integrated Regressors," *Econometric Theory*, 11, 1131–1147. [528]
- Davidson, J. (1994), *Stochastic Limit Theory*, Oxford: Oxford University Press. [531]
- Deng, A. (2014), "Understanding Spurious Regression in Financial Economics," *Journal of Financial Econometrics*, 12, 122–150. [529,532]
- Deo, R. S. (2000), "Spectral Tests of the Martingale Hypothesis Under Conditional Heteroskedasticity," *Journal of Econometrics*, 99, 291–315. [536]
- Elliott, G., Müller, U., and Watson, M. W. (2015), "Nearly Optimal Tests When a Nuisance Parameter is Present Under the Null Hypothesis," *Econometrica*, 83, 771–811. [529,532]
- Ferson, W., Sarkissian, S., and Simin, T. (2003a), "Spurious Regression in Financial Economics?" *Journal of Finance*, 58, 1393–1414. [529,532]
- (2003b), "Is Stock Return Predictability Spurious?" *Journal of Investment Management*, 1, 10–19. [529]
- Georgiev, I., Harvey, D. I., Leybourne, S. J. and Taylor, A. M. R. (2018), "Testing for Parameter Instability in Predictive Regression Models," *Journal of Econometrics*, 204, 101–108. [536]
- Gonçalves, S., and Kilian, L. (2004), "Bootstrapping Autoregressions With Conditional Heteroskedasticity of Unknown Form," *Journal of Econometrics*, 123, 89–120. [529,536]
- Gospodinov, N. (2009), "A New Look at the Forward Premium Puzzle," *Journal of Financial Econometrics*, 7, 312–338. [530]
- Hansen, B. E. (2000), "Testing for Structural Change in Conditional Models," *Journal of Econometrics*, 97, 93–115. [529,535]
- Jansson, M., and Moreira, M. J. (2006), "Optimal Inference in Regression Models With Nearly Integrated Regressors," *Econometrica*, 74, 681–714. [529,532]
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992), "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?" *Journal of Econometrics*, 54, 159–178. [529]
- LePage, R., Podgórski, K., and Ryznar, M. (1997), "Strong and Conditional Invariance Principles for Samples Attracted to Stable Laws," *Probability Theory and Related Fields*, 108, 281–298. [535]
- Leybourne, S. J., and McCabe, B. P. M. (1994), "A Consistent Test for a Unit Root," *Journal of Business and Economic Statistics*, 12, 157–167. [529,533]
- Müller, U. (2005), "Size and Power of Tests of Stationarity in Highly Autocorrelated Time Series," *Journal of Econometrics*, 128, 195–213. [529]
- Paye, B. S., and Timmermann, A. (2006), "Instability of Return Prediction Models," *Journal of Empirical Finance*, 13, 274–315. [530]
- Phillips, P. C. B. (2014), "On Confidence Intervals for Autoregressive Roots and Predictive Regression," *Econometrica*, 82, 1177–1195. [528,532]
- Rossi, B. (2005), "Testing Long Horizon Predictive Ability With High Persistence, and the Meese-Rogoff Puzzle," *International Economic Review*, 46, 61–92. [528]
- Rubinshtein, B. Z. (1996), "A Central Limit Theorem for Conditional Distributions," in *Convergence in Ergodic Theory and Probability*, eds. V. Bergelson, P. March, and J. Rosenblatt, Berlin: De Gruyter, pp. 373–380. [536]
- Shin, Y. (1994), "A Residual-Based Test of the Null of Cointegration Against the Alternative of No Cointegration," *Econometric Theory*, 10, 91–115. [529,533]