

# A Bootstrap Stationarity Test for Predictive Regression Invalidation\*

Iliyan Georgiev<sup>a</sup>, David I. Harvey<sup>b</sup>, Stephen J. Leybourne<sup>b</sup>  
and A.M. Robert Taylor<sup>c</sup>

<sup>a</sup>Department of Statistical Sciences, University of Bologna

<sup>b</sup>School of Economics, University of Nottingham

<sup>c</sup>Essex Business School, University of Essex

## Abstract

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 with the result that both the finite sample and asymptotic size of the predictability tests can be significantly inflated, with the potential therefore to spuriously indicate predictability. 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 heteroskedasticity 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. This entails demonstrating that the asymptotic distribution of the bootstrap statistic, conditional on the data, is the same (to first-order) as the asymptotic 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 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 important applications beyond the present context. An illustration is given by re-examining the results relating to U.S. stock returns data in Campbell and Yogo (2006).

**Keywords:** Predictive regression; Granger causality; persistence; stationarity test; fixed regressor wild bootstrap; conditional distribution.

**JEL Classification:** C12, C32.

---

\*We are grateful to the Editor, Todd Clark, an anonymous Co-Editor and three anonymous referees for their helpful and constructive comments. We particularly thank one of the referees for suggesting the examples relating to latent predictors discussed in section 2. Correspondence to: Robert Taylor, Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. Email: [rtaylor@essex.ac.uk](mailto:rtaylor@essex.ac.uk).

# 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 behaviour akin to a (near) unit root autoregressive process, whilst 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) and, hence, a m.d.s.; the alternative is that it is predictable from  $x_{t-1}$  (i.e. that  $x_{t-1}$  Granger causes  $y_t$ ). Cavanagh *et al.* (1995) [CES] show that when the innovation driving  $x_t$  is correlated with  $y_t$  (as is often thought to be case in practice; e.g., 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 towards 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 (2006) [CY] consider a point optimal variant of the  $t$ -test and employ confidence belts. Phillips (2014) proposes a 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 (2015) [BD] consider variable addition and instrumental variable (IV) methods to correct test size. Near-optimal PR tests can also be found in Elliott *et al.* (2015) and Jansson and Moreira (2006).

A misspecified PR of  $y_t$  on  $x_{t-1}$  (with non-zero 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. Here, and in the extreme case where  $y_t$  and the past of  $x_t$  are independent, such that  $x_{t-1}$  is an invalid predictor variable, 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 *et al.* (2003a,b) and Deng (2014). More generally, where  $x_t$  and  $z_t$  are correlated such that  $y_t$  and the past of  $x_t$  are no longer independent, 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  $x_{t-1}$ . Specifically, where  $z_t$  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 even in the case of Granger non-causality<sup>1</sup> of  $y_t$  by the process  $\{x_t\}$ ,  $x_{t-1}$  would not be a spurious predictor variable; nevertheless, the optimal linear predictor of  $y_t$  would involve further variables among the lags of  $y_t$  and  $x_{t-1}$ . Where  $z_t$  is manifest, the optimal linear predictor for  $y_t$  would involve the past of  $z_t$ . 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 modelling 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.* (1992) [KPSS]. Although originally designed to detect pure unit root behaviour in re-

---

<sup>1</sup>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.

gression residuals, Müller (2005) shows these tests also reject when near unit root behaviour is present, making them well-suited to the testing scenario of this paper.

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 analogues 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 non-stationary volatility and/or conditional heteroskedasticity in their innovations, it is also important for our proposed testing procedure to be (asymptotically) robust to these effects. We therefore use a heteroskedasticity-robust variant of the fixed regressor bootstrap along the lines proposed in Hansen (2000). This uses a wild bootstrap scheme to generate bootstrap analogues 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 asymptotic distribution of the bootstrap statistic, conditional on the data, is the same (to first-order) as the asymptotic null distribution of the statistic computed on the original data, conditional on the posited predictor variable. Our proof corrects a long-standing error in the bootstrap literature arising from Hansen (2000) who incorrectly suggests, in the context of a different test statistic, that for strongly persistent regressors the validity of the fixed regressor bootstrap follows from the coincidence of the unconditional limiting distribution of the original statistic with that of the limiting distribution of the bootstrap statistic conditional of the data. We show that this coincidence does not occur for strongly persistent regressors. The results we provide have wider applicability to other scenarios where a fixed regressor bootstrap is used with (near-) integrated regressors.

The paper is organised 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. Generalisations 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  $\xrightarrow{p}$  for weak convergence and convergence in probability, respectively. For a vector,  $x$ ,  $\|x\| := (x'x)^{1/2}$ , the Euclidean norm. Finally,  $\mathcal{D}^k := D_k[0, 1]$  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 in order to examine the asymptotic local power of the test procedures

we discuss, we parameterise  $\beta_x$  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 non-zero,  $y_t$  is a persistent, but local-to-noise process.<sup>2</sup>

Our interest lies in examining the behaviour 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 unincluded  $z_{t-1}$  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 re-written 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 behaviour, such as has been considered in, for example, Paye and Timmermann (2006) and Cai *et al.* (2015), or where the data on  $x_t$  are observed with a strongly persistent measurement error driven by relatively low variance innovations.

---

<sup>2</sup>Notice that an observationally equivalent formulation of the model can be obtained by treating  $\beta_x$  and  $\beta_z$  as fixed constants but parameterising 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.

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  $3 \times 3$  non-stochastic 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 non-stochastic, 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 heteroskedasticity and non-stationary unconditional volatility are obtained as special cases with  $D_t = I_3$  (constant unconditional variance, hence only conditional heteroskedasticity) and  $\sigma_t = I_3$  (so  $\Omega_{t|t-1} = \Omega_t = \Omega(t/T)$ , only unconditional non-stationary volatility), respectively.<sup>3</sup> 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 behaviour 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 also a (local-to-) unit root autoregressive process and its innovations  $\epsilon_{z_t}^\lambda = \epsilon_{z_t} - \lambda \beta_z^{-1} \epsilon_{x_t}$  are such that  $[\epsilon_{x_t}, \epsilon_{z_t}^\lambda, \epsilon_{y_t}]'$  satisfies Assumption 1, upon a redefinition of the matrix  $H$ . In particular, if  $\beta_z \neq 0$ , then it is possible to

---

<sup>3</sup>The assumption that  $E(e_t e_t') = I_3$  made in part (b)(i) and the parameterisation of the unconditionally homoskedastic case by  $D_t = I_3$  are without loss of generality, by non-identification considerations.

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 taking this value into consideration.

**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 (with respect to symmetric quadratic loss) [BLP] 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  $y_{t-1}$  (and, in general, also  $x_{t-1}$  even if  $\beta_x = 0$ , as some of the predictive power available from  $z_{t-1}$  would be picked up by  $x_{t-1}$ ).

**Remark 4.** For transparency, the structure in (1)- (3) is exposted 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  $\{z_t\}$ , unincluded 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):

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

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 behaviour 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 to 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 heteroskedastic 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 homoskedasticity. 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 homoskedastic case  $D_t = I_3$  reduces to

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

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 standardised analogues  $B_{\eta c, u}(r) := \omega_{uu}^{-1/2} M_{\eta c, u}(r)$  and their demeaned counterparts  $\bar{B}_{\eta c, u}(r) := B_{\eta c, u}(r) - \int_0^1 B_{\eta c, u}(s)$ .

### 3 Asymptotic Behaviour 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:

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

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

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

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 favour of predictability of  $y_t$  by  $x_{t-1}$  will be encountered when  $y_t$  is actually predictable by  $z_{t-1}$  (cf. Ferson *et al.*, 2003a,b, and Deng, 2014, for related results under non-localized  $\beta_z$ ). 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

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

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 favour 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 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 *et al.* (2015) and Jansson and Moreira (2006); see the useful recent summaries provided in BD and Cai *et al.* (2015). The issues we discuss in this paper 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 non-stationary 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 2,000 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 behaviour 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  $H_x$ ). 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 focussed 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 predictor  $z_{t-1}$ , but the price paid for employing such an IV is that when a true predictor  $x_{t-1}$  is high persistence, 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  $\beta_z \neq 0$ . Non-rejection 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  $y_t$ , signalling the invalidity of PR tests based on  $x_{t-1}$ . Formally, then, we wish to test the null hypothesis

that  $\beta_z = 0$ , i.e.  $H_u, H_x$ , against the alternative that  $\beta_z \neq 0$ , i.e.  $H_z, H_{xz}$ , 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

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

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

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

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

$$y_t = \alpha_y + \beta_x x_{t-1} + r_{t-1} + \epsilon_{yt} \quad (10)$$

where  $r_t = r_{t-1} + u_t$ , initialised 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 mis-specification 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 mis-specification 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

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

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 standardised heteroskedastic 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  $h_{21}$  (cf. Remark 2). In fact, as  $M_{\eta z} = \omega_{xz}\omega_{xx}^{-1}M_{\eta x} + \omega_{z|x}B_{\eta 2}$  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

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

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 heteroskedasticity 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  $HD_t e_t$  and with the conditions  $\sum_{i=0}^{\infty} i |\theta_i| < \infty$  and  $\sum_{i=0}^{\infty} \theta_i \neq 0$  satisfied. Under homoskedasticity, 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. In this case, it may be shown that the limiting results given in Theorem 2 above and in Theorems 3-5 below continue to hold provided we replace (9) in the calculation of  $S$  with the augmented variant

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

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  $\boldsymbol{\tau}' \mathbf{f}_t$ , 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 below, 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. In order to account for the dependence of inference on any unconditional heteroskedasticity 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 behaviour 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 behaviour of  $x_t$ . For each bootstrap sample, these would then be used to calculate a bootstrap analogue of  $S$ , which should then reflect the behaviour 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 re-sampling 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  $e_{xt}$  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  $y_t^* := \hat{\epsilon}_t w_t$ , where  $w_t$ ,  $t = 1, \dots, T$ , is an  $IID N(0, 1)$  sequence independent of the data and  $\hat{\epsilon}_t$  are the residuals from either (9) or (12).

(ii) Calculate the fixed regressor wild bootstrap analogue of  $S$ ,

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

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

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

(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 favour 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 heteroskedasticity present in the original innovations; this follows because, conditionally on  $\hat{\epsilon}_t$ ,  $y_t^*$  is independent over time with zero mean and variance  $\hat{\epsilon}_t^2$ .

**Remark 12.** By definition, the residuals  $\hat{\epsilon}_t$  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). This is because the  $\hat{\epsilon}_t$  used to construct  $y_t^*$  are free of any effects arising from correlation between  $\epsilon_{xt}$  and  $\epsilon_{yt}$ , or from any weak dependence in  $\epsilon_{xt}$ .

**Remark 13.** Although  $\hat{\epsilon}_t$  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 non-random 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 sized 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, e.g., 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 as a test conditional on  $x$ .

The results we present differ from those given in 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, *ibidem*). The results given in this section show

that any such claim about unconditional inference is not correct, at least for the non-empty 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 in Georgiev *et al.* (2016).

Theorem 2 is based on the invariance principle given in (4). Conditional and bootstrap analogues of that theorem can be based on a *conditional* joint invariance principle for the original and the bootstrap data. In order to obtain this result, we will need to 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 i.i.d. 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 i.i.d. 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 section 3 of Gonçalves and Kilian (2004). It would be interesting, although beyond the scope of our paper, 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 below 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  $\tilde{\epsilon}_{tb} := w_t \tilde{e}_{Tt}$  ( $t = 1, \dots, T$ ), 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  $T \rightarrow \infty$ :*

$$\left( T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \epsilon_t, T^{-1} \sum_{t=1}^T \sum_{s=1}^{t-1} \epsilon_{xs} [\epsilon_{yt}, \epsilon_{zt}] \right) \Big| x \xrightarrow{w} \left( M_\eta(r), \int_0^1 M_{\eta x}(s) d[M_{\eta y}(s), M_{\eta z}(s)] \right) \Big| B_1,$$

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

$$\left( T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} [e_{1t}, \tilde{\epsilon}_{tb}], T^{-1} \sum_{t=1}^T \sum_{s=1}^{t-1} \epsilon_{xs} \tilde{\epsilon}_{tb} \right) \Big| x, y, z \xrightarrow{w} \left( B_1(r), \tilde{B}_\eta(r), \int_0^1 M_{\eta x}(s) d\tilde{B}_\eta(s) \right) \Big| B_1,$$

$r \in [0, 1]$ , 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

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

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 behaviour of  $S$  conditional on  $x$ , and of  $S^*$ , its bootstrap analogue 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

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

given that the random cdf of  $\int_0^1 F(r, c_x)^2 dr \Big| B_1$  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 non-random limit are compared; here, in contrast,  $P_x(S \leq u) \xrightarrow{w} P(\int_0^1 F(r, c_x)^2 dr \leq u | B_1)$  with a non-degenerate random limit.

In Corollary 1 below 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,  $P_T^*|x \xrightarrow{w_p} U[0, 1]$  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 homoskedastic 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 non-monotonicity 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.<sup>4</sup>

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 favour 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 mis-specification 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.

The Supplementary Appendix to this paper contains asymptotic power simulation re-

---

<sup>4</sup>We note that  $S$  is not LBI when we allow correlation between  $\epsilon_{yt}$  and  $\epsilon_{zt}$  so this anomalous behaviour is perhaps not entirely surprising.



sults 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.<sup>5</sup> We also consider the IV-based test of BD that combines fractional and sine function instruments, denoted  $IV_{comb}$ , 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 heteroskedasticity 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  $e_t \sim IID N(0, I_3)$ . Figure 2 reports the finite sample analogues of Figure 1, i.e. rejection frequencies of nominal 0.10-level (two-sided for  $t_u$ ,  $Q$  and  $IV_{comb}$ ) 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 behaviour 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

---

<sup>5</sup>We are grateful to Campbell and Yogo for making their Gauss code available for these two procedures.

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  $IV_{comb}$ ; 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.

Lastly we examine the impact of unconditional heteroskedasticity in the DGP on the size of  $S_B$  and  $IV_{comb}$  when the error processes are subject to a single break in volatility.<sup>6</sup> Specifically, we again simulate the DGP (1)-(3) for  $T = 200$  with  $g_x = g_z = 0$ ,  $e_t \sim IID N(0, I_3)$ , but setting  $d_{it} = 1(t \leq \lfloor \tau T \rfloor) + \sigma_i 1(t > \lfloor \tau T \rfloor)$  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  $IV_{comb}$ ). 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 the large volatility changes that we consider.<sup>7</sup> The  $IV_{comb}$  test also displays a good degree of robustness to heteroskedasticity, 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  $g_z = 50$ , i.e. 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  $d_{2t} = 1(t \leq \lfloor \tau T \rfloor) + \sigma_2 1(t > \lfloor \tau T \rfloor)$ . It is clear that the presence of (unconditional) heteroskedasticity can have a substantial

---

<sup>6</sup>We do not consider  $t_u$  and  $Q$  here since these procedures are not robust to heteroskedastic errors.

<sup>7</sup>We also simulated the finite sample size of  $S_B$  under a variety of conditionally heteroskedastic 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.

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 heteroskedasticity. 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 heteroskedastic.

## 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 analyse the time series behaviour 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  $s_y^2$ .<sup>8</sup> 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 behaviour (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). We conclude that, in accordance with the findings of CY and BD, these possible predictors all display (to differing degrees) strongly persistent behaviour. 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  $IV_{comb}$  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

---

<sup>8</sup>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.

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  $IV_{comb}$  test in Panel C, significant predictability at the 0.10 level is again (as with  $Q$ ) signalled for monthly  $r_3$  and monthly  $y - r_1$ , but also signalled 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 paper 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 unincluded 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. In order to again allow for an unknown degree of persistence in the putative (and latent) predictors, and to allow for both conditional and unconditional heteroskedasticity 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 non-stationary 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

re-consideration 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 paper encourages further research in this area, developing additional within- and out-of-sample diagnostic procedures for PRs.

## References

- Awad, A. (1981). Conditional central limit theorems for martingales and reversed martingales. *Sankhya* 43 A, 100–106.
- Bansal, R. and A. Yaron (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance* 59, 1481–1509.
- Boswijk P., G. Cavaliere, A. Rahbek and A.M.R. Taylor (2016). Inference on co-integration parameters in heteroskedastic vector autoregressions. *Journal of Econometrics*, 192, 64–85.
- Breitung, J. and M. Demetrescu (2015). Instrumental variable and variable addition based inference in predictive regressions. *Journal of Econometrics* 187, 358–375.
- Cai, Z., Y. Wang and Y. Wang (2015). Testing instability in a predictive regression model with nonstationary regressors. *Econometric Theory* 31, 953–980.
- Campbell, J.Y. and R.J. Shiller, (1988). Stock prices, earnings, and expected dividends. *Journal of Finance* 43, 661–676.
- Campbell, J.Y. and M. Yogo (2006). Efficient tests of stock return predictability. *Journal of Financial Economics* 81, 27–60.
- Cavaliere, G. and I. Georgiev (2017). Bootstrap inference under random distributional limits. *Research Institute for Econometrics Discussion Paper* 1-17.

- Cavaliere, G., A. Rahbek and A.M.R. Taylor (2010). Testing for co-integration in vector autoregressions with non-stationary volatility. *Journal of Econometrics* 158, 7–24.
- Cavaliere, G., and A.M.R. Taylor (2005). Stationarity Tests under Time-Varying Second Moments. *Econometric Theory* 21, 1112–1129.
- Cavanagh, C.L., G. Elliott and J.H. Stock (1995). Inference in models with nearly integrated regressors. *Econometric Theory* 11, 1131–1147.
- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford University Press.
- Deng, A. (2014). Understanding spurious regression in financial economics. *Journal of Financial Econometrics* 12, 122–150.
- Deo, R.S. (2000). Spectral tests of the martingale hypothesis under conditional heteroskedasticity. *Journal of Econometrics* 99, 291–315.
- Elliott, G., U. Müller and M.W. Watson (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica* 83, 771–811.
- Ferson, W., S. Sarkissian and T. Simin (2003a). Spurious regression in financial economics? *Journal of Finance* 58, 1393–1414.
- Ferson, W., S. Sarkissian and T. Simin (2003b). Is stock return predictability spurious? *Journal of Investment Management* 1, 10–19.
- Georgiev, I., D.I. Harvey, S.J. Leybourne and A.M.R. Taylor (2016). Testing for Parameter Instability in Predictive Regression Models. Mimeo.
- Gonçalves, S. and L. Kilian (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics* 123, 89–120.
- Gospodinov, N. (2009). A new look at the forward premium puzzle. *Journal of Financial Econometrics* 7, 312–338.
- Hansen, B.E. (2000). Testing for structural change in conditional models. *Journal of Econometrics* 97, 93–115.

- Jansson, M. and M.J. Moreira (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica* 74, 681–714.
- Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (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.
- LePage, R., K. Podgórski, and M. Ryznar (1997). Strong and conditional invariance principles for samples attracted to stable laws. *Probability Theory and Related Fields* 108, 281-298.
- Leybourne, S. J. and B.P.M. McCabe (1994). A consistent test for a unit root. *Journal of Business and Economic Statistics* 12, 157–167.
- Müller, U. (2005). Size and power of tests of stationarity in highly autocorrelated time series. *Journal of Econometrics* 128, 195–213.
- Paye, B.S. and A. Timmermann (2006). Instability of return prediction models. *Journal of Empirical Finance* 13, 274–315.
- Phillips, P.C.B. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica* 82, 1177–1195.
- Rossi, B. (2005). Testing long horizon predictive ability with high persistence, and the Meese-Rogoff Puzzle. *International Economic Review* 46, 61–92.
- Rubshtein B.Z. (1996). A central limit theorem for conditional distributions. In Bergelson V., P. March, J. Rosenblatt (eds.), *Convergence in Ergodic Theory and Probability*, De Gruyter, Berlin.
- Shin, Y. (1994). A residual-based test of the null of cointegration against the alternative of no cointegration. *Econometric Theory* 10, 91–115.



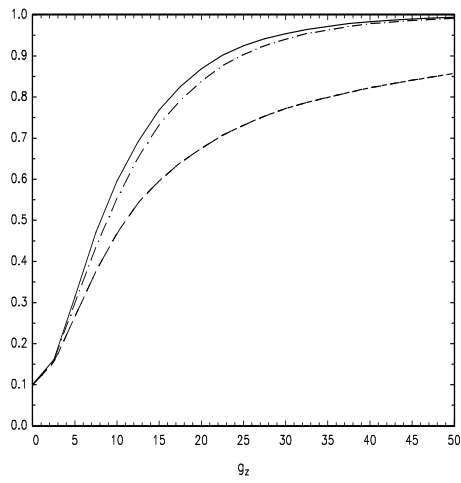
Table 1. Finite sample size of  $S_B$  and  $IV_{comb}$  under volatility shifts:  
 $T = 200, g_x = g_z = 0, d_{it} = 1(t \leq \lfloor \tau T \rfloor) + \sigma_i 1(t > \lfloor \tau T \rfloor), i = 1, 3$

$\sigma_1$	$\sigma_3$	$c_x = 0$				$c_x = 5$				$c_x = 10$			
		$\tau = 0.3$		$\tau = 0.7$		$\tau = 0.3$		$\tau = 0.7$		$\tau = 0.3$		$\tau = 0.7$	
		$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$
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

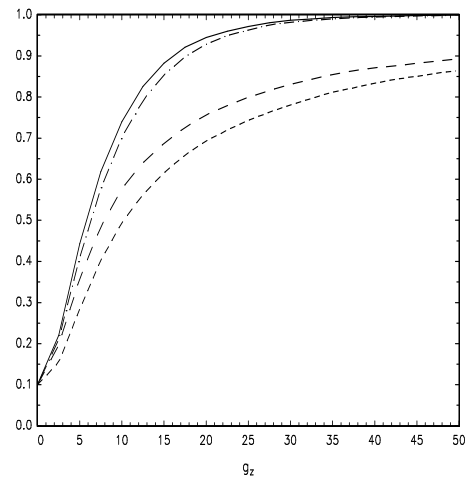
Table 2. Application to U.S. Equity Indices

Series	Obs.	Predictor	$S$	$p$ -val.	$KPSS$	$p$ -val.	$IV_{comb}$	$p$ -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
		$r_3$	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
Quarterly	204	$d - p$	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
		$r_3$	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	*
Monthly	612	$d - p$	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
		$r_3$	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	*

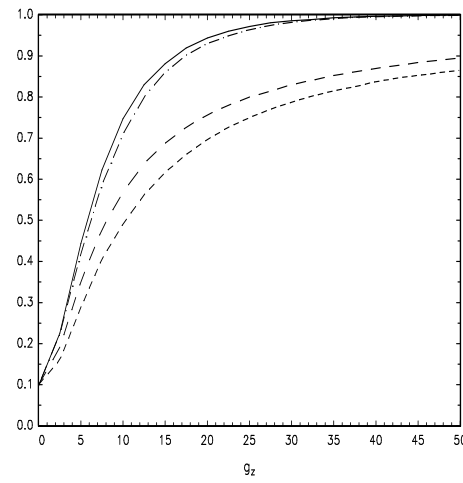
**Notes:** Returns are for the annual S&P 500 index and the annual, quarterly, and monthly CRSP value-weighted index. The predictor variables are the log dividend-price ratio  $d - p$ , the log earnings-price ratio  $e - p$ , the three-month T-bill rate  $r_3$ , and the long-short yield spread  $y - r_1$ . 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.



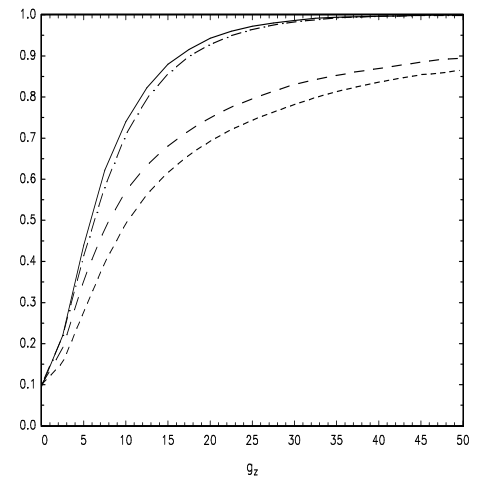
(a)  $c = 0, \sigma_{xy} = 0, \sigma_{zy} = 0$



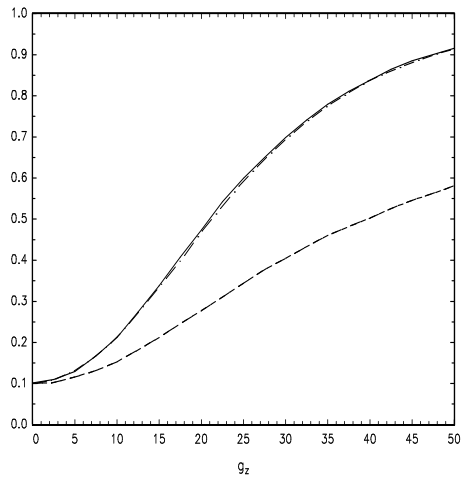
(b)  $c = 0, \sigma_{xy} = -0.7, \sigma_{zy} = 0$



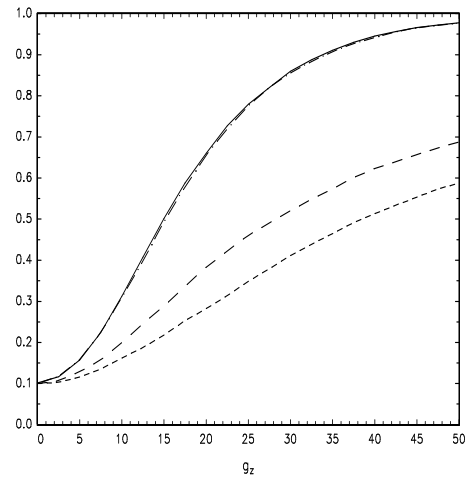
(c)  $c = 0, \sigma_{xy} = -0.7, \sigma_{zy} = -0.7$



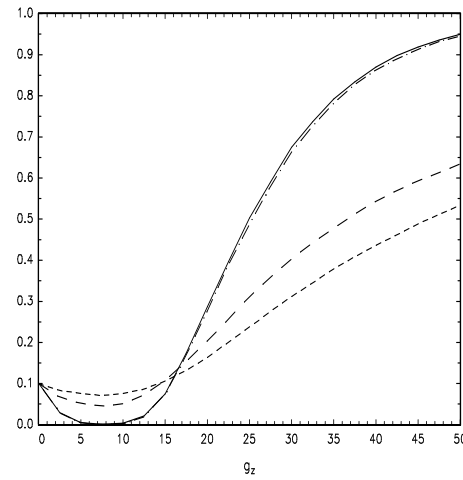
(d)  $c = 0, \sigma_{xy} = -0.7, \sigma_{zy} = 0.7$



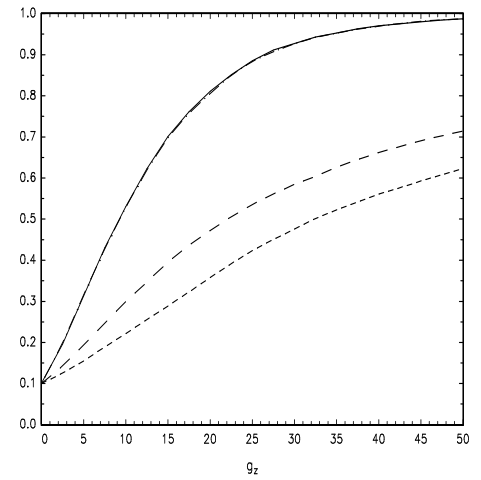
(e)  $c = 10, \sigma_{xy} = 0, \sigma_{zy} = 0$



(f)  $c = 10, \sigma_{xy} = -0.7, \sigma_{zy} = 0$



(g)  $c = 10, \sigma_{xy} = -0.7, \sigma_{zy} = -0.7$



(h)  $c = 10, \sigma_{xy} = -0.7, \sigma_{zy} = 0.7$

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

$S$ :  $-\cdot-$ ,  $S_B$ :  $—$ ,  $t_u$ :  $- - -$ ,  $Q$ :  $- -$

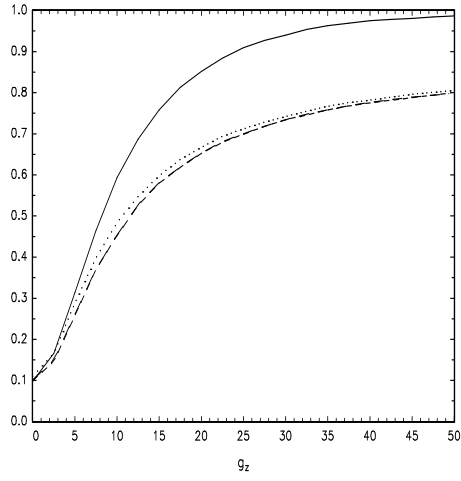
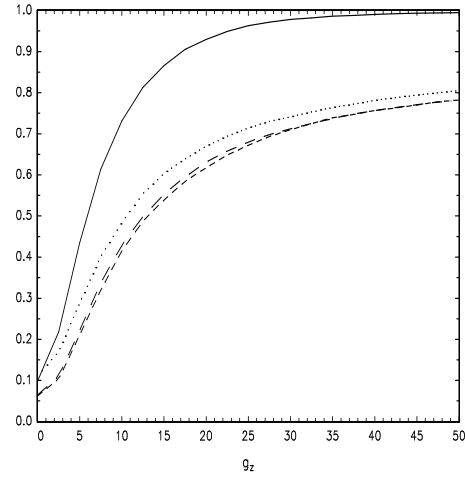
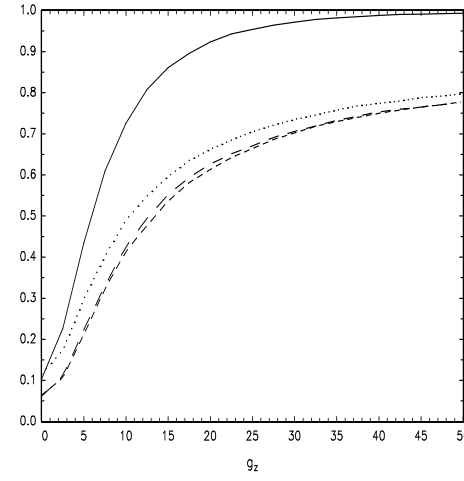
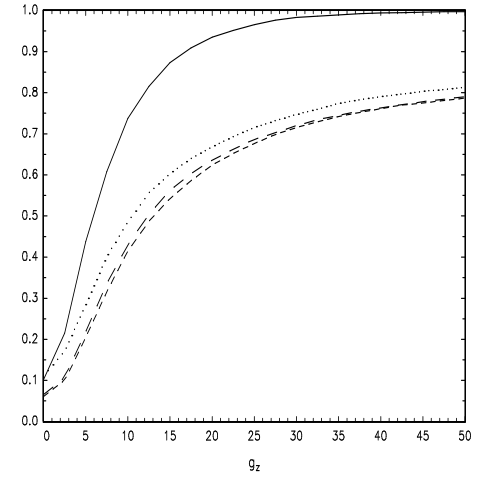
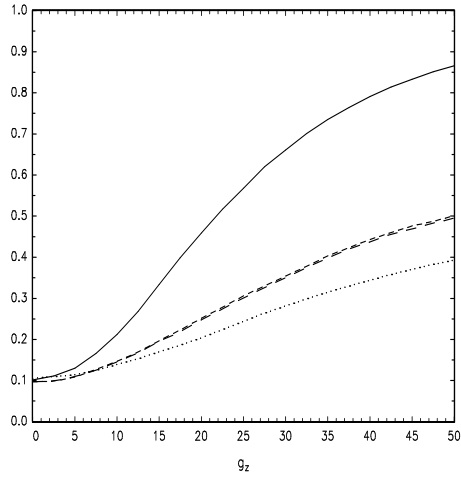
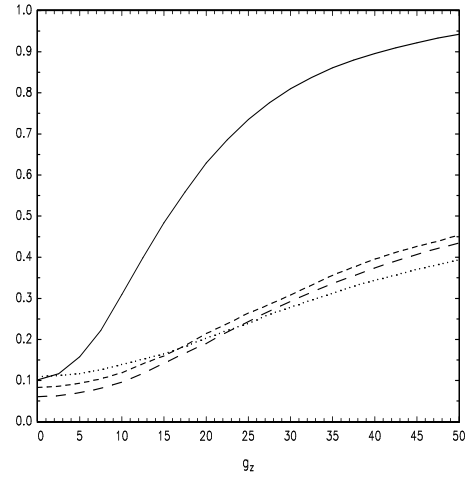
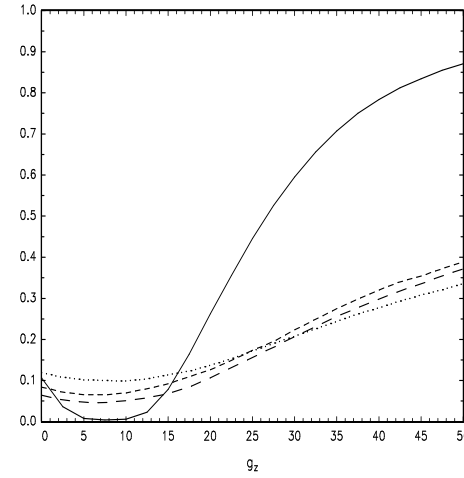
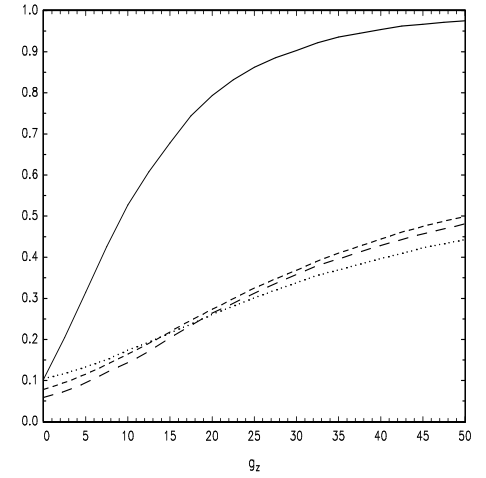
(a)  $c = 0, \sigma_{xy} = 0, \sigma_{zy} = 0$ (b)  $c = 0, \sigma_{xy} = -0.7, \sigma_{zy} = 0$ (c)  $c = 0, \sigma_{xy} = -0.7, \sigma_{zy} = -0.7$ (d)  $c = 0, \sigma_{xy} = -0.7, \sigma_{zy} = 0.7$ (e)  $c = 10, \sigma_{xy} = 0, \sigma_{zy} = 0$ (f)  $c = 10, \sigma_{xy} = -0.7, \sigma_{zy} = 0$ (g)  $c = 10, \sigma_{xy} = -0.7, \sigma_{zy} = -0.7$ (h)  $c = 10, \sigma_{xy} = -0.7, \sigma_{zy} = 0.7$ 

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

$S_B$ : —,  $t_u$ : - - -,  $Q$ : - · -,  $IV_{comb}$ : ···

# Supplementary Online Appendix

to

A Bootstrap Stationarity Test for Predictive Regression Invalidity

by

I. Georgiev, D.I. Harvey, S.J. Leybourne and A.M.R. Taylor

Date: May 15, 2017

## S.1 Introduction

This supplement contains additional Monte Carlo results and proofs for our paper “A Bootstrap Stationarity Test for Predictive Regression Invalidity.” Equation references (S. $n$ ) for  $n \geq 1$  refer to equations in this supplement and other equation references are to the main paper.

The supplement is organised as follows. Additional Monte Carlo simulation results are reported in section S.2. Section S.3 provides mathematical proofs for the large sample results given in the main paper. All additional references are included at the end of the supplement.

## S.2 Additional Monte Carlo Results

Figure S.1 reports asymptotic simulation results for the same tests and DGP settings as for Figure 1, but replacing  $c = 0$  and  $c = 10$  with  $c = 5$  and  $c = 20$ , respectively. Figure S.2 reports similar results, but allowing for  $c_x \neq c_z$ . Figures S.3-S.6 report, for various combinations of  $c_x$  and  $c_z$ , results for  $\sigma_{xz} = \pm 0.5$ , with the magnitudes of  $\sigma_{xy}$  and  $\sigma_{zy}$  reduced in some cases to ensure  $\Omega$  remains positive definite.

Tables S.1 and S.2 report finite sample results for the same tests and DGP settings as for Table 1, but with  $g_z = 25$  and  $g_z = 50$ , with  $c_z = c_x$  and additionally allowing for a volatility break in  $\epsilon_{zt}$  via  $d_{2t} = 1(t \leq \lfloor \tau T \rfloor) + \sigma_2 1(t > \lfloor \tau T \rfloor)$ .

### S.3 Mathematical Proofs

We start with some preliminaries. First, we set  $s_{x,0} = s_{z,0} = 0$  throughout the Appendix, without loss of generality under our assumptions. Second, for centred variables we introduce the notation  $\dot{y}_t := y_t - \bar{y}$ ,  $\dot{x}_t := x_t - \bar{x}_{-1}$  and  $\Delta \dot{x}_t := \Delta x_t - \overline{\Delta x}$ , where  $\bar{y} := T^{-1} \sum_{t=1}^T y_t$ ,  $\bar{x}_{-1} := T^{-1} \sum_{t=0}^{T-1} x_t$  and  $\overline{\Delta x} := T^{-1} \sum_{t=1}^T \Delta x_t$ .

Third, we will repeatedly use the following result, which holds under Assumption 1 by virtue of Lemma A.1 of Boswijk *et al.* (2016),

$$T^{-1} \sum_{t=1}^T \epsilon_t \epsilon_t' \xrightarrow{p} \Omega_\eta = H \left[ \int_0^1 \text{diag}\{d_1^2(r), d_2^2(r), d_3^2(r)\} dr \right] H' = H \text{diag}\{f_1, f_2, f_3\} H' = H F H' \quad (\text{S.1})$$

where  $\text{diag}\{v\}$  denotes a diagonal matrix with  $v$  on the main diagonal.

Fourth, we will also use the functional Orstein-Uhlenbeck convergence

$$T^{-1/2} \begin{bmatrix} x_{\lfloor Tr \rfloor} \\ z_{\lfloor Tr \rfloor} \end{bmatrix} \xrightarrow{w} \int_0^r \begin{bmatrix} e^{-(r-s)c_x} dM_{\eta x}(s) \\ e^{-(r-s)c_z} dM_{\eta z}(s) \end{bmatrix} = \begin{bmatrix} M_{\eta c, x}(r) \\ M_{\eta c, z}(r) \end{bmatrix} =: M_{\eta c}(r), \quad r \in [0, 1], \quad (\text{S.2})$$

and the associated convergence to stochastic integrals

$$T^{-1} \sum_{t=1}^T \begin{bmatrix} x_{t-1} \\ z_{t-1} \end{bmatrix} [\epsilon_t', \Delta x_t, \Delta z_t] \xrightarrow{w} \int_0^1 M_{\eta c}(s) d[M_\eta(s)', M_{\eta c}(s)']. \quad (\text{S.3})$$

These obtain from (4) by routine arguments using a standard approximation of the exponential function, partial summation and integration, and the continuous mapping theorem [CMT].

**Proof of Theorem 1:** We may set  $\alpha_y$ ,  $\alpha_x$  and  $\alpha_z$  to zero, without loss of generality.

First write  $t_u$  as

$$t_u = \frac{T^{-1} \sum_{t=1}^T \dot{x}_{t-1} y_t}{\sqrt{s_y^2 T^{-2} \sum_{t=1}^T \dot{x}_{t-1}^2}}.$$

Then, we can write

$$\begin{aligned} T^{-1} \sum_{t=1}^T \dot{x}_{t-1} y_t &= g_x T^{-2} \sum_{t=1}^T \dot{x}_{t-1} x_{t-1} + g_z T^{-2} \sum_{t=1}^T \dot{x}_{t-1} z_{t-1} + T^{-1} \sum_{t=1}^T \dot{x}_{t-1} \epsilon_{yt} \\ &\xrightarrow{w} g_x \int_0^1 \bar{M}_{\eta c, x}(r)^2 + g_z \int_0^1 \bar{M}_{\eta c, x}(r) M_{\eta c, z}(r) + \int_0^1 \bar{M}_{\eta c, x}(r) dM_{\eta y}(r) \end{aligned}$$

and  $T^{-2} \sum_{t=1}^T \dot{x}_{t-1}^2 \xrightarrow{w} \int_0^1 \bar{M}_{\eta c, x}(r)^2$  using (S.2), (S.3) and the CMT. Also,

$$\begin{aligned} s_y^2 &= T^{-1} \sum_{t=1}^T \dot{y}_t^2 - T^{-1} \frac{\{T^{-1} \sum_{t=1}^T \dot{x}_{t-1} y_t\}^2}{T^{-2} \sum_{t=1}^T \dot{x}_{t-1}^2} + o_p(1) = T^{-1} \sum_{t=1}^T y_t^2 - \bar{y}^2 + o_p(1) \\ &= T^{-1} \sum_{t=1}^T (g_x T^{-1} x_{t-1} + g_z T^{-1} z_{t-1} + \epsilon_{yt})^2 \\ &\quad - \left\{ T^{-1} \sum_{t=1}^T (g_x T^{-1} x_{t-1} + g_z T^{-1} z_{t-1} + \epsilon_{yt}) \right\}^2 + o_p(1) \\ &= T^{-1} \sum_{t=1}^T \epsilon_{yt}^2 + o_p(1) \xrightarrow{p} \omega_{yy} \end{aligned}$$

by (S.1). Consequently, by the CMT,

$$t_u \xrightarrow{w} \frac{g_x \int_0^1 \bar{M}_{\eta c, x}(r)^2 + g_z \int_0^1 \bar{M}_{\eta c, x}(r) M_{\eta c, z}(r) + \int_0^1 \bar{M}_{\eta c, x}(r) dM_{\eta y}(r)}{\sqrt{\omega_{yy} \int_0^1 \bar{M}_{\eta c, x}(r)^2}}.$$

It follows from the previous discussion of  $\sum_{t=1}^T \dot{x}_{t-1} y_t$  and  $\sum_{t=1}^T \dot{x}_{t-1}^2$  that

$$T \hat{\beta}_x \xrightarrow{w} \frac{g_x \int_0^1 \bar{M}_{\eta c, x}(r)^2 + g_z \int_0^1 \bar{M}_{\eta c, x}(r) M_{\eta c, z}(r) + \int_0^1 \bar{M}_{\eta c, x}(r) dM_{\eta y}(r)}{\int_0^1 \bar{M}_{\eta c, x}(r)^2}.$$

Also,

$$T(\hat{\rho}_x - \rho_x) = \frac{T^{-1} \sum_{t=1}^T \dot{x}_{t-1} \epsilon_{xt}}{T^{-2} \sum_{t=1}^T \dot{x}_{t-1}^2} \xrightarrow{w} \frac{\int_0^1 \bar{M}_{\eta c, x}(r) dM_{\eta x}(r)}{\int_0^1 \bar{M}_{\eta c, x}(r)^2}$$

since  $T^{-1} \sum_{t=1}^T \dot{x}_{t-1} \epsilon_{xt} \xrightarrow{w} \int_0^1 \bar{M}_{\eta c, x}(r) dM_{\eta x}(r)$  using (S.2), (S.3) and the CMT. Now

$$\begin{aligned} \hat{\epsilon}_{xt} &= x_t - \bar{x} - \hat{\rho}_x \dot{x}_{t-1} \\ &= \rho_x x_{t-1} + \epsilon_{xt} - \rho_x \bar{x}_{-1} - \bar{\epsilon}_x - \hat{\rho}_x \dot{x}_{t-1} \\ &= -(\hat{\rho}_x - \rho_x) \dot{x}_{t-1} + \epsilon_{xt} - \bar{\epsilon}_x \end{aligned}$$

giving

$$\begin{aligned}
s_x^2 &= T^{-1} \sum_{t=1}^T \{ -(\hat{\rho}_x - \rho_x) \dot{x}_{t-1} + \epsilon_{xt} - \bar{\epsilon}_x \}^2 + o_p(1) \\
&= (\hat{\rho}_x - \rho_x)^2 T^{-1} \sum_{t=1}^T \dot{x}_{t-1}^2 + T^{-1} \sum_{t=1}^T (\epsilon_{xt} - \bar{\epsilon}_x)^2 \\
&\quad - 2(\hat{\rho}_x - \rho_x) T^{-1} \sum_{t=1}^T \dot{x}_{t-1} (\epsilon_{xt} - \bar{\epsilon}_x) + o_p(1) \\
&= T^{-1} \sum_{t=1}^T \epsilon_{xt}^2 + o_p(1) \xrightarrow{p} \omega_{xx}
\end{aligned}$$

by (S.1), and

$$\begin{aligned}
s_{xy} &= T^{-1} \sum_{t=1}^T \hat{\epsilon}_{xt} \hat{\epsilon}_{yt} + o_p(1) \\
&= T^{-1} \sum_{t=1}^T \{ -(\hat{\rho}_x - \rho_x) \dot{x}_{t-1} + \epsilon_{xt} - \bar{\epsilon}_x \} \{ \beta_x \dot{x}_{t-1} + \beta_z \dot{z}_{t-1} + (\epsilon_{yt} - \bar{\epsilon}_y) - \hat{\beta}_x \dot{x}_{t-1} \} + o_p(1) \\
&= T^{-1} \sum_{t=1}^T \epsilon_{xt} \epsilon_{yt} + o_p(1) \xrightarrow{p} \omega_{xy}
\end{aligned}$$

using (S.1).

So, using the limit of  $s_y^2$  from the discussion of  $t_u$ , we find that

$$\begin{aligned}
Q &= \frac{T \hat{\beta}_x - (s_{xy}/s_x^2) T (\hat{\rho}_x - \rho_x)}{\sqrt{s_y^2 \{ 1 - (s_{xy}^2/s_y^2 s_x^2) \} / T^{-2} \sum_{t=1}^T (x_{t-1} - \bar{x}_{-1})^2}} \\
&\xrightarrow{w} \frac{g_x \int_0^1 \bar{M}_{\eta c, x}(r)^2 + g_z \int_0^1 \bar{M}_{\eta c, x}(r) M_{\eta c, z}(r) + \int_0^1 \bar{M}_{\eta c, x}(r) dM_{\eta y}(r) - \omega_{xy} \omega_{xx}^{-1} \int_0^1 \bar{M}_{\eta c, x}(r) dM_{\eta x}(r)}{\sqrt{(\omega_{yy} - \omega_{xy}^2/\omega_{xx}) \int_0^1 \bar{M}_{\eta c, x}(r)^2}} \\
&= \frac{g_x \int_0^1 \bar{M}_{\eta c, x}(r)^2 + g_z \int_0^1 \bar{M}_{\eta c, x}(r) M_{\eta c, z}(r) + \int_0^1 \bar{M}_{\eta c, x}(r) d\{M_{\eta y}(r) - \omega_{xy} \omega_{xx}^{-1} M_{\eta x}(r)\}}{\sqrt{\omega_{y|x} \int_0^1 \bar{M}_{\eta c, x}(r)^2}}
\end{aligned}$$

■

**Proof of Theorem 2:** We may set  $\alpha_y$ ,  $\alpha_x$  and  $\alpha_z$  to zero, and  $g_x$  to  $-ch_{11}^{-1}h_{31}$ , without loss of generality, since the  $\hat{\epsilon}_t$  are invariant to these parameters. Let  $y_t^x := y_t - h_{11}^{-1}h_{31}\Delta x_t$ ,  $\dot{y}_t^x := \dot{y}_t - h_{11}^{-1}h_{33}\Delta \dot{x}_t$  and  $\epsilon_{yt}^x := \epsilon_{yt} - h_{31}d_{1t}e_{1t} = h_{32}d_{2t}e_{2t} + h_{33}d_{3t}e_{3t}$ . For later reference



we first observe that

$$\begin{aligned} T^{-1} \sum_{t=1}^T \dot{x}_{t-1} y_t^x &= T^{-1} \sum_{t=1}^T \dot{x}_{t-1} \epsilon_{yt}^x + g_z T^{-1} \sum_{t=1}^T \dot{x}_{t-1} z_{t-1} \\ &\xrightarrow{w} \int_0^1 \bar{M}_{\eta c, x}(r) d\{\omega_{y|x}^{1/2} B_{\eta, y|x}(r)\} + g_z \int_0^1 \bar{M}_{\eta c, x}(r) M_{\eta c, z}(r) \end{aligned} \quad (\text{S.4})$$

using (S.2), (S.3) and the CMT, with  $\omega_{y|x} = h_{32}^2 f_2 + h_{33}^2 f_3$  and  $B_{\eta, y|x}(r) = \omega_{y|x}^{-1/2} \{h_{32} f_2^{1/2} B_{\eta 2}(r) + h_{33} f_3^{1/2} B_{\eta 3}(r)\}$ .

Next, consider the limit of the partial sum process for  $\hat{e}_t$ , which we write as

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{e}_t = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \dot{y}_t - \left[ T^{-3/2} \sum_{t=1}^{\lfloor Tr \rfloor} \dot{x}_{t-1} \quad T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \Delta \dot{x}_t \right] N_T \hat{\beta} \quad (\text{S.5})$$

with  $N_T := \text{diag}\{1, T\}$  and

$$N_T \hat{\beta} := \begin{bmatrix} T^{-2} \sum_{t=1}^T \dot{x}_{t-1}^2 & T^{-1} \sum_{t=1}^T \dot{x}_{t-1} \Delta x_t \\ T^{-2} \sum_{t=1}^T \dot{x}_{t-1} \Delta x_t & T^{-1} \sum_{t=1}^T (\Delta \dot{x}_t)^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-1} \sum_{t=1}^T \dot{x}_{t-1} y_t \\ T^{-1} \sum_{t=1}^T \Delta \dot{x}_t y_t \end{bmatrix}.$$

Before passing to the limit in (S.5), we focus on  $N_T \hat{\beta}$ . It holds that

$$N_T \hat{\beta} = \Delta_T^{-1} \begin{bmatrix} T^{-1} \sum_{t=1}^T (\Delta \dot{x}_t)^2 & -T^{-1} \sum_{t=1}^T \dot{x}_{t-1} \Delta x_t \\ o_p(1) & T^{-2} \sum_{t=1}^T \dot{x}_{t-1}^2 \end{bmatrix} \begin{bmatrix} T^{-1} \sum_{t=1}^T \dot{x}_{t-1} y_t \\ T^{-1} \sum_{t=1}^T \Delta \dot{x}_t y_t \end{bmatrix}, \quad (\text{S.6})$$

where  $\Delta_T := T^{-3} \{\sum_{t=1}^T \dot{x}_{t-1}^2 \sum_{t=1}^T (\Delta \dot{x}_t)^2 - (\sum_{t=1}^T \dot{x}_{t-1} \Delta x_t)^2\} = T^{-3} \sum_{t=1}^T \dot{x}_{t-1}^2 \sum_{t=1}^T (\Delta \dot{x}_t)^2 + o_p(T^{-3})$  because  $\sum_{t=1}^T \dot{x}_{t-1} \Delta x_t = O_p(T)$  by (S.2) and (S.3). Further, as also  $\sum_{t=1}^T \dot{x}_{t-1} y_t = O_p(T)$  by the proof of Theorem 1, it holds that

$$\begin{aligned} N_T \hat{\beta} &= \Delta_T^{-1} \begin{bmatrix} T^{-2} \{\sum_{t=1}^T \dot{x}_{t-1} y_t \sum_{t=1}^T (\Delta \dot{x}_t)^2 - \sum_{t=1}^T \dot{x}_{t-1} \Delta x_t \sum_{t=1}^T \Delta \dot{x}_t y_t\} \\ T^{-3} \sum_{t=1}^T \dot{x}_{t-1}^2 \sum_{t=1}^T \Delta \dot{x}_t y_t + o_p(1) \end{bmatrix} \\ &= \Delta_T^{-1} \begin{bmatrix} T^{-2} \{\sum_{t=1}^T \dot{x}_{t-1} y_t^x \sum_{t=1}^T (\Delta \dot{x}_t)^2 - \sum_{t=1}^T \dot{x}_{t-1} \Delta x_t \sum_{t=1}^T \Delta \dot{x}_t y_t^x\} \\ T^{-3} \frac{h_{31}}{h_{11}} \sum_{t=1}^T \dot{x}_{t-1}^2 \sum_{t=1}^T (\Delta \dot{x}_t)^2 + T^{-3} \sum_{t=1}^T \dot{x}_{t-1}^2 \sum_{t=1}^T \Delta \dot{x}_t y_t^x + o_p(1) \end{bmatrix} \\ &= \Delta_T^{-1} \begin{bmatrix} T^{-2} \sum_{t=1}^T \dot{x}_{t-1} y_t^x \sum_{t=1}^T (\Delta \dot{x}_t)^2 + o_p(1) \\ T^{-3} \frac{h_{31}}{h_{11}} \sum_{t=1}^T \dot{x}_{t-1}^2 \sum_{t=1}^T (\Delta \dot{x}_t)^2 + o_p(1) \end{bmatrix} \end{aligned} \quad (\text{S.7})$$

because  $\sum_{t=1}^T \Delta \dot{x}_t y_t^x = \sum_{t=1}^T \Delta x_t \epsilon_{yt}^x + g_z T^{-1} \sum_{t=1}^T \Delta x_t z_{t-1} - T^{-1} (x_T - x_1) \{\sum_{t=1}^T \epsilon_{yt}^x + g_z T^{-1} \sum_{t=1}^T z_{t-1}\} = o_p(T)$  given that (i)  $\sum_{t=1}^T \Delta x_t \epsilon_{yt}^x = \sum_{t=1}^T \epsilon_{xt} \epsilon_{yt}^x - c T^{-1} \sum_{t=1}^T x_{t-1} \epsilon_{yt}^x =$

$o_p(T)$  using (S.1) and the convergence  $T^{-1} \sum_{t=1}^T x_{t-1} \epsilon_{yt}^x \xrightarrow{w} \int_0^1 M_{\eta c, x}(s) d\{\omega_{y|x}^{1/2} B_{\eta, y|x}(s)\}$  implied by (S.3), (ii)  $T^{-1} \sum_{t=1}^T \Delta x_t z_{t-1} \xrightarrow{w} \int_0^1 M_{\eta c, z}(r) dM_{\eta c, x}(r)$  as a consequence of (S.3), (iii)  $T^{-1/2}(x_T - x_1) \xrightarrow{w} M_{\eta c, x}(1)$  by (S.2) and the CMT, (iv)  $T^{-1/2} \sum_{t=1}^T \epsilon_{yt}^x \xrightarrow{w} \omega_{y|x}^{1/2} B_{\eta, y|x}(1)$ , and (v)  $T^{-3/2} \sum_{t=1}^T z_{t-1} \xrightarrow{w} \int_0^1 M_{\eta c, z}(s)$  by (S.2) and the CMT. Finally,

$$N_T \hat{\beta} = \left[ (T^{-1} \sum_{t=1}^T \dot{x}_{t-1}^2)^{-1} \sum_{t=1}^T \dot{x}_{t-1} y_t^x \quad h_{31}^{-1} h_{33} \right]' + o_p(1) \quad (\text{S.8})$$

because  $T^{-1} \sum_{t=1}^T (\Delta \dot{x}_t)^2 = T^{-1} \sum_{t=1}^T \epsilon_{tx}^2 - 2c_x T^{-2} \sum_{t=1}^T \epsilon_{tx} x_{t-1} + T^{-3} c_x^2 \sum_{t=1}^T x_{t-1}^2 - T^{-2} (x_T - x_1)^2 = T^{-1} \sum_{t=1}^T \epsilon_{tx}^2 + o_p(1) \xrightarrow{p} \omega_{xx}$  by (S.1), so  $T^{-1} \sum_{t=1}^T (\Delta \dot{x}_t)^2$  is bounded away from zero in  $P$ -probability.

Given (S.8), (S.5) simplifies to

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{e}_t = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \dot{y}_t^x - \frac{\sum_{t=1}^{\lfloor Tr \rfloor} \dot{x}_{t-1} y_t^x}{T^{-1} \sum_{t=1}^{\lfloor Tr \rfloor} \dot{x}_{t-1}^2} T^{-3/2} \sum_{t=1}^{\lfloor Tr \rfloor} \dot{x}_{t-1} + \rho_T(r), \quad (\text{S.9})$$

where

$$\begin{aligned} \sum_{t=1}^{\lfloor Tr \rfloor} \dot{y}_t^x &= \sum_{t=1}^{\lfloor Tr \rfloor} \epsilon_{yt}^x + T^{-1} g_z \sum_{t=1}^{\lfloor Tr \rfloor} z_{t-1} - \frac{\lfloor Tr \rfloor - 1}{T} \left\{ \sum_{t=1}^T \epsilon_{yt}^x + T^{-1} g_z \sum_{t=1}^T z_{t-1} \right\} \\ &\xrightarrow{w} \omega_{y|x}^{1/2} (B_{\eta, y|x}(r) - r B_{\eta, y|x}(1)) + g_z \left( \int_0^r M_{\eta c, z}(s) - r \int_0^r M_{\eta c, z} \right) \end{aligned}$$

on  $\mathcal{D}$ , and  $\rho_T(r) = o_p(1) T^{-3/2} \sum_{t=1}^{\lfloor Tr \rfloor} \dot{x}_{t-1} + o_p(1) T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \Delta \dot{x}_t$  is such that

$$\sup_{r \in [0, 1]} |\rho_T(r)| \leq o_p(1) \sup_{r \in [0, 1]} |T^{-3/2} \sum_{t=1}^{\lfloor Tr \rfloor} \dot{x}_{t-1}| + o_p(1) T^{-1/2} \sup_{t=0, \dots, T} |x_t| = o_p(1) \quad (\text{S.10})$$

because  $\sup_{r \in [0, 1]} |T^{-3/2} \sum_{t=1}^{\lfloor Tr \rfloor} \dot{x}_{t-1}| \xrightarrow{w} \sup_{r \in [0, 1]} \left| \int_0^r \bar{M}_{\eta c, x}(s) \right|$  and  $T^{-1/2} \sup_{t=0, \dots, T} |x_t| \xrightarrow{w} \sup_{r \in [0, 1]} |M_{\eta c, x}(r)|$  by the CMT. Therefore, using also (S.4) and the CMT again,

$$\begin{aligned} T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{e}_t &\xrightarrow{w} \omega_{y|x}^{1/2} \left\{ B_{\eta, y|x}(r) - r B_{\eta, y|x}(1) - \frac{\int_0^1 \bar{M}_{\eta c, x}(s) dB_{\eta, y|x}(s)}{\int_0^1 \bar{M}_{\eta c, x}^2(s)} \int_0^r \bar{M}_{\eta c, x}(s) \right\} \\ &\quad + g_z \left\{ \int_0^r M_{\eta c, z}(s) - r \int_0^1 M_{\eta c, z}(s) - \frac{\int_0^1 \bar{M}_{\eta c, x}(s) M_{\eta c, z}(s)}{\int_0^1 \bar{M}_{\eta c, x}^2(s)} \int_0^r \bar{M}_{\eta c, x}(s) \right\} \\ &= \omega_{y|x}^{1/2} \{ F(r, c_x) + g_z G(r, c_x, c_z) \} \end{aligned}$$

on  $\mathcal{D}$ .

Next, using the previously established order of magnitude results, we have that,

$$\begin{aligned}
\sum_{t=1}^T \hat{\epsilon}_t^2 &= \sum_{t=1}^T \dot{y}_t^2 - \left[ T^{-1} \sum_{t=1}^T \dot{x}_{t-1} y_t \quad \sum_{t=1}^T \Delta \dot{x}_t y_t \right] N_T \hat{\beta} \\
&= \sum_{t=1}^T \dot{y}_t^2 - h_{31}^{-1} h_{33} \sum_{t=1}^T \Delta \dot{x}_t y_t - \sum_{t=1}^T \dot{x}_{t-1} y_t \left( \sum_{t=1}^T \dot{x}_{t-1}^2 \right)^{-1} \sum_{t=1}^T \dot{x}_{t-1} y_t^x + o_p(T) \\
&= \sum_{t=1}^T \dot{y}_t^2 - h_{31}^{-2} h_{33}^2 \sum_{t=1}^T (\Delta \dot{x}_t)^2 - h_{31}^{-1} h_{33} \sum_{t=1}^T \Delta \dot{x}_t y_t^x + o_p(T) \\
&= \sum_{t=1}^T (\dot{y}_t^x)^2 + h_{31}^{-1} h_{33} \sum_{t=1}^T y_t^x \Delta \dot{x}_t + o_p(T) \\
&= \sum_{t=1}^T (\epsilon_{yt}^x)^2 - 2T^{-1} g_z \sum_{t=1}^T z_{t-1} \epsilon_{yt} + T^{-2} g_z^2 \sum_{t=1}^T z_{t-1}^2 + o_p(T) = \sum_{t=1}^T (\epsilon_{yt}^x)^2 + o_p(T),
\end{aligned} \tag{S.11}$$

where  $T^{-1} \sum_{t=1}^T (\epsilon_{yt}^x)^2 \xrightarrow{p} h_{32}^2 f_2 + h_{33}^2 f_3 = \omega_{y|x}$  by (S.1). Consequently,

$$s^2 \xrightarrow{p} \omega_{y|x}, \tag{S.12}$$

and by the CMT,  $S \xrightarrow{w} \int_0^1 \{F(r, c_x) + g_z G(r, c_x, c_z)\}^2 dr$ . ■

Before proceeding to the proof of Theorem 5, we first define some additional notation related to the conditional convergence modes used in the remainder of the Appendix. For weak convergence of random measures induced by conditioning, i.e., of the form  $(\cdot)|x \xrightarrow{w} (\circ)|B_1$  and  $(\blacktriangle)|x, y, z \xrightarrow{w} (\triangle)|B_1$ , we write  $(\cdot) \xrightarrow{w_x} (\circ)|B_1$  and  $(\cdot) \xrightarrow{w^*} (\triangle)|B_1$  respectively, the definitions being  $E\{f(\cdot)|x\} \xrightarrow{w} E\{f(\circ)|B_1\}$  and  $E\{g(\blacktriangle)|x, y, z\} \xrightarrow{w} E\{g(\triangle)|B_1\}$  for all bounded continuous real functions  $f$  and  $g$ , where  $\cdot$ ,  $\circ$ ,  $\blacktriangle$  and  $\triangle$  are placeholders for random elements. We say that the  $w_x$  and  $w^*$  convergence are joint if  $(E\{f(\cdot)|x\}, E\{g(\blacktriangle)|x, y, z\})' \xrightarrow{w} (E\{f(\circ)|B_1\}, E\{g(\triangle)|B_1\})'$  for the same class of functions  $f, g$ . This is distinct from the two  $w_x$  modes of convergence,  $(\cdot) \xrightarrow{w_x} (\circ)|B_1$  and  $(\blacktriangle) \xrightarrow{w_x} (\triangle)|B_1$ , being joint, where  $E\{h(\cdot, \blacktriangle)|x\} \xrightarrow{w} E\{h(\circ, \triangle)|B_1\}$  should hold for bounded continuous  $h$  (and similarly, for  $w^*$ ). We write  $(\cdot)_T = O_p^x(1)$  to denote that for every  $\varepsilon > 0$  there exists a  $C > 0$  such that  $P(P(\|(\cdot)_T\| > C|x) > \varepsilon) < \varepsilon$ , and  $(\cdot)_T = o_p^x(1)$  if  $(\cdot)_T \xrightarrow{w_x} 0$ , where  $\|\cdot\|$  is a norm (for random processes, the uniform norm). The corresponding notation  $O_p^*(1)$  and  $o_p^*(1)$  is introduced similarly for conditioning on the data.

In Theorem 5 we now establish a homoskedastic joint conditional and bootstrap invariance principle.

**Theorem 5.** *Define the partial sums  $U_{ti} := T^{-1/2} \sum_{s=1}^t e_{is}$  ( $i = 1, 2, 3$ ),  $U_t := [U_{t1}, U_{t2}, U_{t3}]'$  and  $U_{tb} := T^{-1/2} \sum_{s=1}^t e_s w_s$ . Moreover, let  $B^\dagger := [B_1^\dagger, B_2^\dagger, B_3^\dagger]'$  denote a standard Brownian motion in  $\mathbb{R}^3$ , independent of  $B$ . Under Assumption 2, the following converge jointly as  $T \rightarrow \infty$ :*

$$U_{[T \cdot]} | x \xrightarrow{w} B | B_1$$

and

$$[U_{[T \cdot]1}, U'_{[T \cdot]b}]' | x, y, z \xrightarrow{w} [B_1, (B^\dagger)']' | B_1$$

in the sense of weak convergence of random measures on  $\mathcal{D}^3$  and  $\mathcal{D}^4$ .

Let  $E_x(\cdot) := E(\cdot | x)$  and  $E^*(\cdot) := E(\cdot | x, y, z)$ . The definition of the joint weak convergence of random measures result established in Theorem 5, is that for all bounded continuous real functions  $f$  and  $g$  on  $\mathcal{D}^3$  and  $\mathcal{D}^4$ , respectively, it holds that

$$\begin{bmatrix} E_x(f(U'_{[T \cdot]})) \\ E^*(g(U_{[T \cdot]1}, U'_{[T \cdot]b})) \end{bmatrix} \xrightarrow{w} \begin{bmatrix} E(f(B') | B_1) \\ E(g(B_1, (B^\dagger)') | B_1) \end{bmatrix}$$

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

**Proof of Theorem 5:** From Theorem 2 of Rubshtein (1996), by extending the argument to the multivariate case, it follows that  $E(f(U_{[T \cdot]2}, U_{[T \cdot]3}) | \mathcal{X}) \xrightarrow{a.s.} E(f(B_2, B_3))$  for continuous bounded real  $f$  on  $\mathcal{D}^2$ . Then, by the bounded and martingale convergence theorems,

$$E_x f(U_{[T \cdot]2}, U_{[T \cdot]3}) \xrightarrow{p} E f(B_2, B_3) \tag{S.13}$$

for these functions  $f$ . As additionally  $U_{[T \cdot]} \xrightarrow{w} B$  in  $\mathcal{D}^3$  (a special case of (4)), from Corollary 4.1 of Crimaldi and Pratelli (2005) it follows that

$$E_x f(U'_{[T \cdot]}) \xrightarrow{w} E(f(B') | B_1) \tag{S.14}$$

for continuous bounded real  $f$  on  $\mathcal{D}^3$ . Here we have used the result that conditioning on  $x$  and  $U_{[T \cdot]1}$  are equivalent.

Next, we note that  $U_{tb}$ , given the data, is a Gaussian process with independent increments, mean zero and variance function  $V_T(r) := Var^*(U_{[Tr]b}) = T^{-1} \sum_{t=1}^{[Tr]} e_t e_t' \xrightarrow{p} r I_3$

( $r \in [0, 1]$ ), by Lemma A.1 of Boswijk *et al.* (2015). As  $V_T$  are component-wise increasing in  $r$  and their point-wise limit is continuous in  $r$ , the convergence of  $V_T$  is uniform in  $r$ , and it follows that

$$E^* f(U'_{[T \cdot]b}) \xrightarrow{p} E f(B^\dagger) \quad (\text{S.15})$$

for continuous bounded real  $f$  on  $\mathcal{D}^3$ . Additionally,  $[U'_{[T \cdot]}, U'_{[T \cdot]b}]' \xrightarrow{w} [B', B^\dagger]'$  on  $\mathcal{D}^6$  by the martingale functional CLT [MFCLT] of Brown (1971), and so from Corollary 4.1 of Crimaldi and Pratelli (2005) it follows further that, for continuous bounded real  $f$  on  $\mathcal{D}^6$ ,

$$E^* f(U'_{[T \cdot]}, U'_{[T \cdot]b}) \xrightarrow{w} E\{f(B', B^\dagger)|B\};$$

here we have used the result that conditioning on  $x, y, z$  and  $U_{[T \cdot]}$  are equivalent. In particular, for  $f$  that do not depend on  $U_{[T \cdot]1}, U_{[T \cdot]2}$ , restricted to  $\mathcal{D}^4$ , the bootstrap counterpart of (S.14) is obtained:

$$E^* f(U_{[T \cdot]1}, U'_{[T \cdot]b}) \xrightarrow{w} E\{f(B_1, B^\dagger)|B\} = E\{f(B_1, B^\dagger)|B_1\}, \quad (\text{S.16})$$

the last equality following by the independence of the components of  $[B', B^\dagger]'$ .

To see that (S.14) and (S.16) are joint, it is sufficient to apply the Cramer-Wald device to obtain

$$aE_x f(U'_{[T \cdot]}) + bE^* g(U_{[T \cdot]1}, U'_{[T \cdot]b}) \xrightarrow{w} E(af(B') + bg(B_1, B^\dagger)|B_1) \quad (\text{S.17})$$

for arbitrary  $a, b \in \mathbb{R}$  and for continuous bounded real  $f$  and  $g$  on  $\mathcal{D}^3$  and  $\mathcal{D}^4$ , respectively. To this end, by Skorokhod's representation theorem applied to the Polish space  $\mathcal{D}^6$ , and since  $[B', B^\dagger]'$  has a.s. continuous sample paths, we can consider a probability space where  $[U_{[T \cdot]}, U'_{[T \cdot]b}]' \rightarrow [B', B^\dagger]'$  a.s. On this probability space, by Corollary 4.4 of Crimaldi and Pratelli (2005), (S.14) and (S.16) hold in probability instead of weakly, and hence, (S.17) holds in probability. Since the distribution of the involved conditional expectations only depends on  $[U'_{[T \cdot]}, U'_{[T \cdot]b}]'$  and  $[B', B^\dagger]'$ , it follows that on general probability spaces (S.17) holds weakly. ■

**Proof of Theorem 3:** Let  $\tilde{U}_{tb} := T^{-1/2} \sum_{s=1}^t \tilde{\epsilon}_{sb}$  be the bootstrap partial sums. Introduce also  $\tilde{\epsilon}_{it} := d_t e_{it}$ ,  $\tilde{U}_{ti} := T^{-1/2} \sum_{s=1}^t \tilde{\epsilon}_{is}$ ,  $\tilde{M}_i(r) := \int_0^r d_i(s) dB_i(s)$  ( $i = 1, 2, 3; r \in [0, 1]$ ),  $\tilde{U}_t := [\tilde{U}_{t1}, \tilde{U}_{t2}, \tilde{U}_{t3}]'$ ,  $\tilde{M} := [\tilde{M}_1, \tilde{M}_2, \tilde{M}_3]'$ . Given that  $\epsilon_t$  is a linear transformation of  $\tilde{\epsilon}_t$ ,

and linear transformations are continuous on the support of the process  $\tilde{M}$ , it suffices to establish that

$$\left( \tilde{U}_{[T\cdot]}, \sum_{t=1}^T \tilde{U}_{t-1,1} [\Delta \tilde{U}_{t2}, \Delta \tilde{U}_{t3}] \right) \xrightarrow{w_{\mathfrak{F}}} \left( \tilde{M}, \int_0^1 \tilde{M}_1(s) d[\tilde{M}_2(s), \tilde{M}_3(s)] \right) \Big|_{B_1} \quad (\text{S.18})$$

jointly with

$$\left( U_{[T\cdot]1}, \tilde{U}_{[T\cdot]b}, \sum_{t=1}^T \tilde{U}_{t-1,1} \Delta \tilde{U}_{tb} \right) \xrightarrow{w^*} \left( B_1, \tilde{B}_\eta, \int_0^1 \tilde{M}_1(s) d\tilde{B}_\eta(s) \right) \Big|_{B_1}. \quad (\text{S.19})$$

We shall prove Theorem 3 in this way.

Notice first that, given the data,  $\tilde{U}_{[T\cdot]b}$  is a Gaussian process with independent increments, mean zero and variance function  $\text{Var}^*(\tilde{U}_{[Tr]b}) = T^{-1} \sum_{t=1}^{\lfloor Tr \rfloor} \tilde{e}_{Tt}^2$ . Under the assumption that  $T^{-1} \sum_{t=1}^{\lfloor Tr \rfloor} \tilde{e}_{Tt}^2 \xrightarrow{p} \int_0^r m^2(s) ds$ ,  $r \in [0, 1]$ , this convergence is uniform in  $r$  because  $T^{-1} \sum_{t=1}^{\lfloor Tr \rfloor} \tilde{e}_{Tt}^2$  are increasing in  $r$  and the limit integral is continuous in  $r$ . This suffices for the conclusion that  $\tilde{U}_{[T\cdot]b}$  given the data (and thus, given  $U_{[T\cdot]}$ ) converges weakly in probability to  $\tilde{B}_\eta$ :

$$E^* g(\tilde{U}_{[T\cdot]b}) \xrightarrow{p} E g(\tilde{B}_\eta) \quad (\text{S.20})$$

for all bounded continuous real  $g$  on  $\mathcal{D}$ , where  $\tilde{B}_\eta$  is a Gaussian process with independent increments, zero mean and variance function  $\int_0^\cdot m^2(s) ds$ . On the other hand, since  $U_{[T\cdot]} \xrightarrow{w} B$  by the MFCLT of Brown (1971), and since  $\mathcal{D}^3 \times \mathcal{D}$  is separable, it follows that  $[U'_{[T\cdot]}, \tilde{U}_{[T\cdot]b}]' \xrightarrow{w} [B', \tilde{B}_\eta]'$  on  $\mathcal{D}^3 \times \mathcal{D}$ , with  $B$  and  $\tilde{B}_\eta$  independent (see Theorem 2.8 of Billingsley (1999)), and also on  $\mathcal{D}^4$ , because the limit process is continuous.

In view of Skorokhod's representation theorem and the a.s. continuity of  $[B', \tilde{B}_\eta]'$ 's sample paths, we may assume in the remainder of the proof that  $[U'_{[T\cdot]}, \tilde{U}_{[T\cdot]b}]'$  and  $[B', \tilde{B}_\eta]'$  are defined on the same probability space (say  $\mathbb{S}$ ), and

$$[U'_{[T\cdot]}, \tilde{U}_{[T\cdot]b}]' \rightarrow [B', \tilde{B}_\eta]' \text{ a.s.} \quad (\text{S.21})$$

By using (S.21) and the distributional properties of  $[U'_{[T\cdot]}, \tilde{U}_{[T\cdot]b}]'$  (though not functional relations with the data and the bootstrap multipliers, which need not be defined on  $\mathbb{S}$ ), we show that on  $\mathbb{S}$  the convergence in (S.18)-(S.19) holds in probability, so in general it holds weakly. To be specific, we write  $\tilde{U}_{ti} = \sum_{s=1}^t d_i(s/t) \Delta U_{ti}$  ( $i = 1, 2, 3$ ), and establish that on

$\mathbb{S}$ ,

$$E_x \phi \left( \tilde{U}'_{[T]\cdot}, \sum_{t=1}^T \tilde{U}_{t-1,1} [\Delta \tilde{U}_{t2}, \Delta \tilde{U}_{t3}] \right) \xrightarrow{p} E \left[ \phi \left( \tilde{M}', \int_0^1 \tilde{M}_1(s) d[\tilde{M}_2(s), \tilde{M}_3(s)] \right) \middle| B_1 \right] \quad (\text{S.22})$$

and

$$E^* \psi \left( U_{[T]\cdot 1}, \tilde{U}_{[T]\cdot b}, \sum_{t=1}^T \tilde{U}_{t-1,1} \Delta \tilde{U}_{tb} \right) \xrightarrow{p} E \left[ \psi \left( B_1, \tilde{B}_\eta, \int_0^1 \tilde{M}_1(s) d\tilde{B}_\eta(s) \right) \middle| B_1 \right] \quad (\text{S.23})$$

for every bounded and continuous real  $\phi$  and  $\psi$  on  $\mathcal{D}^3 \times \mathbb{R}^2$  and  $\mathcal{D}^2 \times \mathbb{R}$ , respectively. On  $\mathbb{S}$ ,  $E_x$  and  $E^*$  denote exclusively  $E(\cdot | U_{[T]\cdot 1})$  and  $E(\cdot | U_{[T]\cdot})$ . In view of (S.13) and (S.20), on  $\mathbb{S}$  we can still invoke

$$E_x f(U_{[T]\cdot 2}, U_{[T]\cdot 3}) \xrightarrow{w} E f(B_2, B_3) \quad \text{and} \quad E^* g(\tilde{U}_{[T]\cdot b}) \xrightarrow{w} E g(\tilde{B}_\eta)$$

for arbitrary bounded and continuous real  $f$  and  $g$  on  $\mathcal{D}^2$  and  $\mathcal{D}$ , respectively, because the distributions of the conditional expectations depend only on the distributions of  $[U'_{[T]\cdot}, \tilde{U}_{[T]\cdot}]'$  and  $[B', \tilde{B}_\eta]'$ . Moreover, in view also of (S.21), by Corollary 4.4 of Crimaldi and Pratelli (2005), it holds on  $\mathbb{S}$  that

$$E_x h(U'_{[T]\cdot}) \xrightarrow{p} E \{h(B') | B_1\} \quad \text{and} \quad E^* g(U_{[T]\cdot 1}, \tilde{U}_{[T]\cdot b}) \xrightarrow{p} E \{g(B_1, \tilde{B}_\eta) | B_1\} \quad (\text{S.24})$$

for arbitrary bounded and continuous real  $h$  and  $g$  on  $\mathcal{D}^3$  and  $\mathcal{D}^2$ .

It is well known that (S.22)-(S.23) cannot be put in the form of (S.24) for any choice of  $h$  and  $g$  because, in general, the stochastic integrals involved are not continuous transformations. Therefore, we resort to their continuous approximations, as is habitually done.

We approximate:

(a)  $\tilde{U}_{[T]\cdot j}$  by  $\xi_{\delta_j}(U_{[T]\cdot j})$  ( $j = 1, 2, 3$ ), where  $\xi_{\delta_j} : \mathcal{D} \rightarrow \mathcal{D}$  are defined by  $\xi_{\delta_j}(X) = X(\cdot) \delta_j(\cdot) - \int_0^\cdot X(s) \delta'_j(s) ds$  and are continuous on the support  $C[0, 1]$  of  $B_j$  for every fixed smooth function  $\delta_j : [0, 1] \rightarrow \mathbb{R}$ . Then, using (S.24) and integration by parts, it follows that

$$\begin{aligned} & E_x m(\xi_{\delta_1}(U_{[T]\cdot 1}), \xi_{\delta_2}(U_{[T]\cdot 2}), \xi_{\delta_3}(U_{[T]\cdot 3})) \xrightarrow{p} E \{m(\xi_{\delta_1}(B_1), \xi_{\delta_2}(B_2), \xi_{\delta_3}(B_3)) | B_1\} \\ & = E \{m(\int_0^\cdot \delta_1(s) dB_1(s), \int_0^\cdot \delta_2(s) dB_2(s), \int_0^\cdot \delta_3(s) dB_3(s)) | B_1\} \end{aligned}$$

and

$$\begin{aligned} & E^* n(U_{[T \cdot]1}, \xi_{\delta_1}(U_{[T \cdot]1}), \tilde{U}_{[T \cdot]b}) \xrightarrow{p} E\{n(B_1, \xi_{\delta_1}(B_1), \tilde{B}_\eta) | B_1\} \\ & = E\{n(B_1, \int_0^\cdot \delta_1(s) dB_1(s), \tilde{B}_\eta) | B_1\}. \end{aligned}$$

for continuous  $m, n : \mathcal{D}^3 \rightarrow \mathbb{R}$ . It then needs to be argued that the integrals involving smooth  $\delta_j$  approximate those involving  $d_j$ , in conditional distribution, such that it also holds that  $E_x m(\tilde{U}_{[T \cdot]}) \xrightarrow{p} E\{m(\tilde{M}) | B_1\}$  and

$$E^* n(U_{[T \cdot]1}, \tilde{U}_{[T \cdot]1}, \tilde{U}_{[T \cdot]b}) \xrightarrow{p} E\{n(B_1, \tilde{M}_1, \tilde{B}_\eta) | B_1\}.$$

(b)  $\int_0^1 \tilde{U}_{[Ts-]1} d\tilde{U}_{[Ts]j}$  ( $j = 2, 3$ ) and  $\int_0^1 \tilde{U}_{[Ts-]1} d\tilde{U}_{[Ts]b}$  by  $\zeta_L(\tilde{U}_{[T \cdot]1}, \tilde{U}_{[T \cdot]j})$  and  $\zeta_L(\tilde{U}_{[T \cdot]1}, \tilde{U}_{[T \cdot]b})$ , where  $\zeta_L : \mathcal{D}^2 \rightarrow \mathbb{R}$  is defined by

$$\zeta_L(X, Y) := X(1)Y(1) - \sum_{i=1}^L Y\left(\frac{i}{L}\right) \left\{ X\left(\frac{i}{L}\right) - X\left(\frac{i-1}{L}\right) \right\} = \int_0^1 X^L(s-) dY(s),$$

with

$$X^L(s) := \sum_{i=1}^L X\left(\frac{i-1}{L}\right) \mathbb{I}\left\{\frac{i-1}{L} \leq s < \frac{i}{L}\right\} + X(1) \mathbb{I}\{s = 1\},$$

and is continuous on the support of  $[\tilde{M}_1, \tilde{M}_j]'$  and  $[\tilde{M}_1, \tilde{B}_\eta]'$  for every  $L \in \mathbb{N}$ . Then, by an appropriate choice of  $m$  and  $n$  above, it follows that

$$E_x \phi\left(\tilde{U}_{[T \cdot]}, \zeta_L(\tilde{U}_{[T \cdot]1}, \tilde{U}_{[T \cdot]2}), \zeta_L(\tilde{U}_{[T \cdot]1}, \tilde{U}_{[T \cdot]3})\right) \xrightarrow{p} E\left[\phi\left(\tilde{M}, \zeta_L(\tilde{M}_1, \tilde{M}_2), \zeta_L(\tilde{M}_1, \tilde{M}_3)\right) \middle| B_1\right]$$

and

$$E^* \psi\left(U_{[T \cdot]1}, \tilde{U}_{[T \cdot]b}, \zeta_L(\tilde{U}_{[T \cdot]1}, \tilde{U}_{[T \cdot]b})\right) \xrightarrow{p} E\left[\psi\left(B_1, \tilde{B}_\eta, \zeta_L(\tilde{M}_1, \tilde{B}_\eta)\right) \middle| B_1\right]$$

for  $\phi$  and  $\psi$  as in (S.22)-(S.23). To complete the proof, it remains to be shown that, as  $L \rightarrow \infty$ ,  $\zeta_L$  approximates the stochastic integrals of interest sufficiently well, again in conditional distribution.

We turn to the accuracy of these approximations introduced previously, starting from point (a) and proceeding in two steps.

(a.1) By partial summation and the mean-value theorem,

$$\max_{r \in [0,1]} |\tilde{U}_{[Tr]j} - \xi_{\delta_j}(U_{[T \cdot]j})(r)| \leq \max_{r \in [0,1]} \left| \frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor rT \rfloor} \{d_j(\frac{t}{T}) - \delta_j(\frac{t}{T})\} \Delta U_{tj} \right| + \frac{1}{2} \max_{r \in [0,1]} |R_T(r)|, \quad (\text{S.25})$$



where  $R_T(r) := T^{-2} \sum_{t=1}^{\lfloor rT \rfloor} U_{t-1,j} \delta_j''(\theta_t/T)$ , with  $\theta_t \in [(t-1)/T, t/T]$ , satisfies

$$\max_{r \in [0,1]} |R_T(r)| \leq T^{-1} \max_{r \in [0,1]} |\delta_j''(r)| \max_{t=1, \dots, T} |U_{tj}| = o_p^x(1)$$

because  $\{\max_{t=1, \dots, T} |U_{tj}|\} |x \rightarrow \max_{[0,1]} |B_j|$  (a.s. for  $j = 1$  and weakly in probability for  $j = 2, 3$ ) by continuity of the sup on the support of  $B_j$ . Moreover, for  $j = 1$  and every  $\lambda > 0$ , by Doob's inequality and the property  $E(\Delta U_{t1} \Delta U_{s1}) = \mathbb{I}\{t = s\}$  (inherited on  $\mathbb{S}$  from the martingale difference property of  $e_{1t}$  and the standardisation  $Ee_{1t}^2 = 1$ ), it holds that

$$\begin{aligned} & P \left\{ P_x \left( \max_{r \in [0,1]} \left| \frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor rT \rfloor} \{d_1(\frac{t}{T}) - \delta_1(\frac{t}{T})\} \Delta U_{t1} \right| \geq \lambda \right) = 0 \right\} \\ &= 1 - P \left( \max_{r \in [0,1]} \left| \frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor rT \rfloor} \{d_1(\frac{t}{T}) - \delta_1(\frac{t}{T})\} \Delta U_{t1} \right| \geq \lambda \right) \\ &\geq 1 - \frac{1}{\lambda^2} E \left( \frac{1}{T^{1/2}} \sum_{t=1}^T \{d_1(\frac{t}{T}) - \delta_1(\frac{t}{T})\} \Delta U_{t1} \right)^2 \\ &= 1 - \frac{1}{\lambda^2 T} \sum_{t=1}^T \{d_1(\frac{t}{T}) - \delta_1(\frac{t}{T})\}^2 \xrightarrow{T \rightarrow \infty} 1 - \frac{1}{\lambda^2} \int_0^1 (d_1 - \delta_1)^2. \end{aligned}$$

Since smooth functions are dense in  $L_2[0, 1]$ , this limit can be made as close to 1 as desired by choosing  $\delta_1$  according to  $\lambda$ . On the other hand, for  $j = 2, 3$ , by using  $E_x(\Delta U_{tj} | \{\Delta U_{sj}\}_{s=1}^{t-1}) = 0$  (inherited on  $\mathbb{S}$  from  $E_x(e_{jt} | \mathcal{F}_{t-1}) = 0$ , which is a distributional property), it follows from the conditional version of Doob's inequality that

$$\begin{aligned} & P_x \left( \max_{r \in [0,1]} \left| \frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor rT \rfloor} \{d_j(\frac{t}{T}) - \delta_j(\frac{t}{T})\} \Delta U_{tj} \right| \geq \lambda \right) \tag{S.26} \\ &\leq \frac{1}{\lambda^2} E_x \left( \frac{1}{T^{1/2}} \sum_{t=1}^T \{d_j(\frac{t}{T}) - \delta_j(\frac{t}{T})\} \Delta U_{tj} \right)^2 = \frac{1}{\lambda^2 T} \sum_{t=1}^T \{d_j(\frac{t}{T}) - \delta_j(\frac{t}{T})\}^2 E_x[(\Delta U_{tj})^2] \end{aligned}$$

and from Markov's inequality that

$$\begin{aligned} P \left( \frac{1}{\lambda^2 T} \sum_{t=1}^T \{d_j(\frac{t}{T}) - \delta_j(\frac{t}{T})\}^2 E_x[(\Delta U_{tj})^2] \geq \lambda \right) &\leq \frac{E[(\Delta U_{1j})^2]}{\lambda^3 T} \sum_{t=1}^T \{d_j(\frac{t}{T}) - \delta_j(\frac{t}{T})\}^2 \\ &\xrightarrow{T \rightarrow \infty} \lambda^{-3} \int_0^1 (d_j - \delta_j)^2, \end{aligned}$$

which can be made as small as desired by the choice of  $\delta_j$ .

(a.2) By the continuous-time version of Doob's inequality,

$$\begin{aligned} P\left(\max_{r \in [0,1]} \left| \int_0^r \{d_j(u-) - \delta_j(u-)\} dB_j(u) \right| \geq \lambda\right) &\leq \frac{1}{\lambda^2} E\left(\int_0^1 \{d_j(u-) - \delta_j(u-)\} dB_j(u)\right)^2 \\ &= \lambda^{-2} \int_0^1 (d_j - \delta_j)^2 \end{aligned}$$

can be made as small as desired by the choice of  $\delta_j$ , as in step (a.1).

We consider next the integral approximations in point (b), starting from the non-bootstrap case. Let  $\Delta_{TL}^j := \sum_{t=1}^T \tilde{U}_{t-1,1} \Delta \tilde{U}_{tj} - \zeta_L(\tilde{U}_{[T \cdot]1}, \tilde{U}_{[T \cdot]j})$ . As  $E_x(\Delta U_{tj} | \{\Delta U_{sj}\}_{s=1}^{t-1}) = 0$  ( $j = 2, 3, t = 1, \dots, T$ ), with  $\{Tl_i\}_{i=0}^L = \{\lfloor \frac{Ti}{L} \rfloor\}_{i=0}^L$  and  $j = 2, 3$  it holds that

$$\begin{aligned} E_x\{\Delta_{TL}^j\}^2 &= E_x\left\{\sum_{i=1}^L \sum_{t=Tl_{i-1}+1}^{Tl_i} (\tilde{U}_{t-1,1} - \tilde{U}_{Tl_{i-1},1}) \Delta \tilde{U}_{tj}\right\}^2 \\ &= T^{-1} \sum_{i=1}^L \sum_{t=Tl_{i-1}+1}^{Tl_i} (\tilde{U}_{t-1,1} - \tilde{U}_{Tl_{i-1},1})^2 d_j^2\left(\frac{t}{T}\right) E_x[(\Delta U_{tj})^2] \\ &\leq \sup_{[0,1]} |d_j^2| \sum_{i=1}^L \max_{t=Tl_{i-1}+1, \dots, Tl_i} (\tilde{U}_{t-1,1} - \tilde{U}_{Tl_{i-1},1})^2 \cdot T^{-1} \sum_{t=Tl_{i-1}+1}^{Tl_i} E_x[(\Delta U_{tj})^2]. \end{aligned}$$

Here, first,  $\tilde{U}_{[T \cdot]1} \xrightarrow{P} \tilde{M}_1$  can be established on  $\mathbb{S}$  by using the approximation of  $\tilde{U}_{[T \cdot]1}$  with  $\xi_{\delta 1}(U_{[T \cdot]1})$  as was previously done, and second,  $\gamma_{Tij} := T^{-1} \sum_{t=Tl_{i-1}+1}^{Tl_i} (\Delta U_{tj})^2$  satisfies  $E_x \gamma_{Tij} \xrightarrow{P} l_i - l_{i-1}$  as  $T \rightarrow \infty$ . Indeed,  $E_x \gamma_{Tij} = \Gamma_{Tij,K}^{\leq} + \Gamma_{Tij,K}^{>}$  for every  $K > 0$ , where

$$\begin{aligned} \Gamma_{Tij,K}^{\leq} &:= E_x \left( T^{-1} \sum_{t=Tl_{i-1}+1}^{Tl_i} (\Delta U_{tj})^2 \mathbb{I}\{(\Delta U_{tj})^2 \leq K\} \right) \\ &\xrightarrow{P} (l_i - l_{i-1}) E[(\Delta U_{t1})^2 \mathbb{I}\{(\Delta U_{t1})^2 \leq K\}] \rightarrow l_i - l_{i-1} \end{aligned}$$

as  $T \rightarrow \infty$  followed by  $K \rightarrow \infty$ , by the bounded and martingale convergence theorems (as  $T \rightarrow \infty$ ) and then the monotone convergence theorem (as  $K \rightarrow \infty$ ), and

$$\Gamma_{Tij,K}^{>} := E_x \left( T^{-1} \sum_{t=Tl_{i-1}+1}^{Tl_i} (\Delta U_{tj})^2 \mathbb{I}\{(\Delta U_{tj})^2 > K\} \right) \xrightarrow{P} 0$$

as  $T \rightarrow \infty$  followed by  $K \rightarrow \infty$ , by Markov's inequality and the uniformly bounded fourth moment of  $\Delta U_{tj}$ . Therefore, by Chebyshev's inequality,  $P_x(|\Delta_{TL}^j| \geq \lambda)$  for every  $\lambda > 0$  is bounded above by  $\lambda^{-2}$  times a r.v. converging in probability to

$$\sup_{[0,1]} |d_j^2| \sum_{i=1}^L \max_{r \in [l_{i-1}, l_i]} |\tilde{M}_1(r) - \tilde{M}_1(l_{i-1})|^2 \cdot (l_i - l_{i-1}).$$

Further, using Doob's sub-martingale inequality,

$$\begin{aligned}
& P \left( \sum_{i=1}^L \max_{r \in [l_{i-1}, l_i]} |\tilde{M}_1(r) - \tilde{M}_1(l_{i-1})|^2 \cdot (l_i - l_{i-1}) \geq \lambda \right) \\
& \leq \sum_{i=1}^L \frac{l_i - l_{i-1}}{\lambda} \text{Var}(\tilde{M}_1(l_i) - \tilde{M}_1(l_{i-1})) = \sum_{i=1}^L \frac{l_i - l_{i-1}}{\lambda} \int_{l_{i-1}}^{l_i} d_1^2(s) ds \\
& \leq \frac{1}{\lambda} \max_{i=1, \dots, L} |l_i - l_{i-1}| \int_0^1 d_1^2(s) ds \rightarrow 0
\end{aligned}$$

as  $L \rightarrow \infty$  for every  $\lambda > 0$ . Hence,

$$\lim_{L \rightarrow \infty} \limsup_{T \rightarrow \infty} P \left( P_x \left( \left| \sum_{t=1}^T \tilde{U}_{t-1,1} \Delta \tilde{U}_{tj} - \zeta_L(\tilde{U}_{[T \cdot]1}, \tilde{U}_{[T \cdot]j}) \right| \geq \lambda \right) \geq \lambda \right) = 0.$$

On the other hand, it also holds that

$$\zeta_L(\tilde{M}_1, \tilde{M}_2) = \int_0^1 \tilde{M}_1^L(s-) d\tilde{M}_j(s) \xrightarrow{P} \int_0^1 \tilde{M}_1(s-) d\tilde{M}_j(s) \text{ as } L \rightarrow \infty$$

because  $\int_0^1 (\tilde{M}_1^L(s) - \tilde{M}_1(s))^2 ds \xrightarrow{P} 0$  as  $L \rightarrow \infty$ .

Regarding bootstrap integrals, the argument is similar except that  $E^*(\Delta \tilde{U}_{tb})^2$  appears instead of  $E_x(\Delta U_{tj})^2$ . Since  $E^*(\Delta \tilde{U}_{tb} \Delta \tilde{U}_{sb}) = 0$  for  $t \neq s$  (inherited on  $\mathbb{S}$  from the independence of  $w_t$ ), it holds that

$$\begin{aligned}
E^* \left\{ \sum_{i=1}^L \sum_{t=TL_{i-1}+1}^{Tl_i} (\tilde{U}_{t-1,1} - \tilde{U}_{TL_{i-1},1}) \Delta \tilde{U}_{tb} \right\}^2 &= T^{-1} \sum_{i=1}^L \sum_{t=TL_{i-1}+1}^{Tl_i} (\tilde{U}_{t-1,1} - \tilde{U}_{TL_{i-1},1})^2 E^*(\Delta \tilde{U}_{tb})^2 \\
&\leq \sum_{i=1}^L \max_{t=TL_{i-1}+1, \dots, Tl_i} (\tilde{U}_{t-1,1} - \tilde{U}_{TL_{i-1},1})^2 \cdot T^{-1} \sum_{t=TL_{i-1}+1}^{Tl_i} E^*(\Delta \tilde{U}_{tb})^2 \\
&\xrightarrow{P} \sum_{i=1}^L \max_{r \in [l_{i-1}, l_i]} |\tilde{M}_1(r) - \tilde{M}_1(l_{i-1})|^2 \int_{l_{i-1}}^{l_i} m^2(s) ds
\end{aligned}$$

as  $T \rightarrow \infty$ , as  $T^{-1} \sum_{t=TL_{i-1}+1}^{Tl_i} E^*(\Delta \tilde{U}_{tb})^2 \xrightarrow{P} \int_{l_{i-1}}^{l_i} m^2(s) ds$  is a distributional property inherited on  $\mathbb{S}$  from  $T^{-1} \sum_{t=TL_{i-1}+1}^{Tl_i} \tilde{e}_{Tt}^2 \xrightarrow{P} \int_{l_{i-1}}^{l_i} m^2(s) ds$ . The rest of the argument proceeds as for non-bootstrap integrals. This completes the proof of the theorem.  $\blacksquare$

We next discuss some implications of Theorem 3 for Orstein-Uhlenbeck limits and

stochastic integrals involving them. With  $s_{x,0} = \alpha_x = 0$ , the standard evaluation

$$\begin{aligned} \max_{r \in [0,1]} \left| x_{[Tr]} - e^{-c_x \frac{[Tr]}{T}} \sum_{i=1}^{[Tr]} e^{c_x \frac{i}{T}} \epsilon_{xi} \right| &\leq \max_{r \in [0,1]} \sum_{i=0}^{[Tr]-1} \left| (1 - c_x/T)^i - e^{-c_x \frac{i}{T}} \right| |\epsilon_{x,[Tr]-i}| \\ &\leq |(1 - c_x/T)^T - e^{-c_x}| \max_{[0,1]} |d_1| \sum_{t=1}^T |e_{1t}| = O(1) \end{aligned}$$

holds for almost all  $x$ , by the ergodic theorem. As  $\sum_{i=1}^{[Tr]} e^{c_x \frac{i}{T}} \epsilon_{xi} = h_{11} \sum_{i=1}^{[Tr]} e^{c_x \frac{i}{T}} d_1(\frac{i}{T}) e_{1i}$ , by applying Theorem 3 with  $e^{c_x(\cdot)} d_1(\cdot)$  in place of  $d_1(\cdot)$ , it follows that

$$T^{-1/2} x_{[T\cdot]} \xrightarrow{w_x} h_{11} e^{-c_x(\cdot)} \int_0^\cdot e^{c_x s} d_1(s) dB_1(s) \Big|_{B_1} = M_{\eta_{c,x}} |B_1,$$

and similarly,  $T^{-1/2} z_{[T\cdot]} \xrightarrow{w_x} M_{\eta_{c,z}} |B_1$ , jointly with the convergence in Theorem 3, by the argument for that theorem.

Regarding stochastic integrals, for  $\tilde{\epsilon}_{it}$  ( $i = 2, 3$ ) introduced in the proof of Theorem 3, we find by partial summation that

$$\left(1 - \frac{c_x}{T}\right) \sum_{t=1}^T s_{x,t-1} \tilde{\epsilon}_{it} = s_{x,T} \sum_{t=1}^T \tilde{\epsilon}_{it} - \sum_{t=1}^T \epsilon_{xt} \sum_{s=1}^{t-1} \tilde{\epsilon}_{is} + \frac{c_x}{T} \sum_{t=1}^T s_{x,t-1} \sum_{s=1}^{t-1} \tilde{\epsilon}_{is} - \sum_{t=1}^T \epsilon_{xt} \tilde{\epsilon}_{it},$$

where the following jointly converge by the CMT, Theorem 3 and the discussion in the previous paragraph:

$$\begin{aligned} T^{-1} s_{x,T} \sum_{t=1}^T \tilde{\epsilon}_{it} &\xrightarrow{w_x} M_{\eta_{c,x}}(1) \tilde{M}_i(1) |B_1 \\ T^{-1} \sum_{t=1}^T \epsilon_{xt} \sum_{s=1}^{t-1} \tilde{\epsilon}_{is} &\xrightarrow{w_x} h_{11} \int_0^1 [d\tilde{M}_1(s)] \tilde{M}_i(s) |B_1 \\ T^{-2} \sum_{t=1}^T s_{x,t-1} \sum_{s=1}^{t-1} \tilde{\epsilon}_{is} &\xrightarrow{w_x} h_{11} \int_0^1 \tilde{M}_1(s) \tilde{M}_i(s) ds |B_1. \end{aligned}$$

Moreover,  $T^{-1} \sum_{t=1}^T \epsilon_{xt} \tilde{\epsilon}_{it} = o_p^x(1)$  by the conditional Chebyshev inequality, as

$$T^{-1} Var_x \left( \sum_{t=1}^T \epsilon_{xt} \tilde{\epsilon}_{it} \right) \leq KT^{-1} \sum_{t=1}^T e_{1t}^2 E_x e_{it}^2 \rightarrow KE(e_{1t}^2 e_{it}^2) \text{ a.s.} \quad (\text{S.27})$$

using the martingale difference property and the ergodic theorem, with  $K := h_{11}^2 \sup_{[0,1]} |d_1^2 d_i^2|$ .

Therefore,

$$\begin{aligned} T^{-1} \sum_{t=1}^T s_{x,t-1} \tilde{\epsilon}_{it} &\xrightarrow{w_{\mathfrak{F}}} \left( M_{\eta_{c,x}}(1) \tilde{M}_i(1) - h_{11} \int_0^1 [d\tilde{M}_1(s)] \tilde{M}_i(s) + c_x h_{11} \int_0^1 \tilde{M}_1(s) \tilde{M}_i(s) ds \right) \Big|_{B_1} \\ &= \int_0^1 \tilde{M}_i(s) dM_{\eta_{c,x}}(s) \Big|_{B_1} \end{aligned}$$

jointly with the convergence in Theorem 3 and its implications. By continuity again, as  $T^{-2} \sum_{t=1}^T s_{x,t-1} z_{t-1} \xrightarrow{w_{\mathfrak{F}}} \int_0^1 M_{\eta_{c,x}}(s) M_{\eta_{c,z}}(s) ds \Big|_{B_1}$  and  $T^{-3/2} \sum_{t=1}^{T-1} s_{x,t} \xrightarrow{w_{\mathfrak{F}}} M_{\eta_{c,x}}(1) \Big|_{B_1}$ , it follows for  $\hat{s}_{x,t} := s_{x,t} - T^{-1} \sum_{i=1}^{T-1} s_{x,i}$  and  $\epsilon_{yt}^x := \epsilon_{yt} - h_{31} d_{1t} e_{1t}$  that

$$\begin{aligned} T^{-1} \sum_{t=1}^T \hat{s}_{x,t-1} y_t^x &= T^{-1} \sum_{t=1}^T \hat{s}_{x,t-1} (\epsilon_{yt}^x + T^{-1} g_z z_{t-1}) \tag{S.28} \\ &\xrightarrow{w_{\mathfrak{F}}} \left\{ \int_0^1 \bar{M}_{\eta_{c,x}}(s) d[\omega_{y|x}^{1/2} B_{\eta,y|x}(s)] + g_z \int_0^1 \bar{M}_{\eta_{c,x}}(s) M_{\eta_{c,z}}(s) ds \right\} \Big|_{B_1}, \end{aligned}$$

if  $g_x = 0$ , where  $B_{\eta,y|x}$  is defined in Theorem 2.

**Proof of Theorem 4:** We again set  $\alpha_y, \alpha_x, \alpha_z$  to zero and  $g_x$  to  $-h_{11}^{-1} h_{31} c_x$ , without loss of generality. Notice for further reference that for a random sequence  $\xi_T$ ,

$$\xi_T \xrightarrow{p} K = \text{const} \quad \text{implies that} \quad \xi_T \xrightarrow{w_{\mathfrak{F}}} K \tag{S.29}$$

because  $\xi_T \xrightarrow{p} K$  implies, for bounded continuous  $f$ , that  $E_x f(\xi_T) \xrightarrow{p} f(K)$ .

From relations (S.9)-(S.10), with  $\xi_T = \sup_{r \in [0,1]} |\rho_T(r)|$ , it follows that

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{e}_t = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{y}_t^x - \frac{\sum_{t=1}^T \hat{x}_{t-1} y_t^x}{T^{-1} \sum_{t=1}^T \hat{x}_{t-1}^2} T^{-3/2} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{x}_{t-1} + o_p^x(1)$$

uniformly in  $r$ . Here, from Theorem 3, the convergence  $T^{-1/2} z_{\lfloor T \cdot \rfloor} \xrightarrow{w_{\mathfrak{F}}} M_{\eta_{c,z}} \Big|_{B_1}$  and the CMT,

$$\begin{aligned} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{y}_t^x &= \sum_{t=1}^{\lfloor Tr \rfloor} \epsilon_{yt}^x + T^{-1} g_z \sum_{t=1}^{\lfloor Tr \rfloor} z_{t-1} - \frac{\lfloor Tr \rfloor - 1}{T} \left\{ \sum_{t=1}^T \epsilon_{yt}^x + T^{-1} g_z \sum_{t=1}^T z_{t-1} \right\} \\ &\xrightarrow{w_{\mathfrak{F}}} \left\{ \omega_{y|x}^{1/2} (B_{\eta,y|x}(r) - r B_{\eta,y|x}(1)) + g_z \left( \int_0^r M_{\eta_{c,z}}(s) ds - r \int_0^1 M_{\eta_{c,z}}(s) ds \right) \right\} \Big|_{B_1} \end{aligned}$$

in  $\mathcal{D}$ , so using also (S.28), the convergence  $T^{-1/2} x_{\lfloor T \cdot \rfloor} \xrightarrow{w_{\mathfrak{F}}} M_{\eta_{c,x}} \Big|_{B_1}$  and the CMT, we have that in  $\mathcal{D}$ ,

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{e}_t \xrightarrow{w_{\mathfrak{F}}} \omega_{y|x}^{1/2} \{F(r, c_x) + g_z G(r, c_x, c_z)\} \Big|_{B_1}.$$

Next, (S.12) and (S.29) with  $\xi_T = s_y^2 \xrightarrow{w_x} \omega_{y|x}$ . Consequently, by the CMT,

$$S \xrightarrow{w_x} \int_0^1 \{F(r, c_x) + g_z G(r, c_x, c_z)\}^2 dr | B_1. \quad (\text{S.30})$$

We proceed to part (b). The bootstrap process  $T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} y_t^*$  is of the form of  $\tilde{U}_{\lfloor T \cdot \rfloor b}$  of Theorem 3, with  $\tilde{e}_{Tt} = \hat{e}_t$  satisfying  $T^{-1} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{e}_t^2 = T^{-1} \sum_{t=1}^{\lfloor Tr \rfloor} (\epsilon_{yt}^x)^2 + o_p(1)$ ,  $r \in [0, 1]$ . Under Assumption 1, using Lemma 3 of Boswijk *et al.* (2015), we conclude that  $T^{-1} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{e}_t^2 \xrightarrow{p} h_{32}^2 \int_0^r d_2^2(s) ds + h_{33}^2 \int_0^r d_3^2(s) ds = \int_0^r m^2(s) ds$  with  $m(s) = \sqrt{h_{32}^2 d_2^2(s) + h_{33}^2 d_3^2(s)}$ . As  $B_\eta^\dagger$  is a Gaussian process with independent increments, mean zero and  $\text{Var}(B_\eta^\dagger(r)) = \omega_{y|x}^{1/2} \int_0^r m^2(s) ds$ , from Theorem 3 and its discussion it follows that

$$\left( U_{\lfloor T \cdot \rfloor 1}, T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} y_t^*, \sum_{t=1}^T \tilde{U}_{t-1,1} y_t^* \right) \xrightarrow{w^*} \left( B_1, B_\eta^\dagger, \int_0^1 \tilde{M}_1(s) d\{\omega_{y|x}^{1/2} B_\eta^\dagger(s)\} \right) \Big| B_1$$

jointly with  $T^{-1/2} x_{\lfloor T \cdot \rfloor} \xrightarrow{w^*} M_{\eta c, x} | B_1$  and (S.30).

Next,

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{\epsilon}_{yt}^* = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} (y_t^* - \bar{y}^*) - T^{-3/2} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{x}_{t-1} \frac{T^{-1} \sum_{t=1}^T \hat{x}_{t-1} y_t^*}{T^{-2} \sum_{t=1}^T \hat{x}_{t-1}^2},$$

where by the CMT, the following converge jointly, and jointly with (S.30):  $T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} (y_t^* - \bar{y}^*) \xrightarrow{w^*} \omega_{y|x}^{1/2} \{B_\eta^\dagger(r) - r B_\eta^\dagger(1)\} | B_1$  in  $\mathcal{D}$ ,  $T^{-3/2} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{x}_{t-1} \xrightarrow{w^*} \int_0^r \bar{M}_{\eta c, x}(s) ds | B_1$  in  $\mathcal{D}$ ,  $T^{-1} \sum_{t=1}^T \hat{x}_{t-1} y_t^* \xrightarrow{w^*} \int_0^1 \bar{M}_{\eta c, x}(s) d\{\omega_{y|x}^{1/2} B_\eta^\dagger(s)\} | B_1$  analogously to (S.28),  $T^{-2} \sum_{t=1}^T \hat{x}_{t-1}^2 \xrightarrow{w^*} \int_0^1 \bar{M}_{\eta c, x}^2(s) ds | B_1$ , and since the two limit processes in  $\mathcal{D}$  are continuous,

$$\begin{aligned} & T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{\epsilon}_{yt}^* \xrightarrow{w^*} \omega_{y|x}^{1/2} \left( B_\eta^\dagger(r) - r B_\eta^\dagger(1) - \int_0^r \bar{M}_{\eta c, x}(s) \{ \int_0^1 \bar{M}_{\eta c, x}^2(s) \}^{-1} \int_0^1 \bar{M}_{\eta c, x}(s) dB_\eta^\dagger(s) \right) \Big| B_1 \\ &= \omega_{y|x}^{1/2} \left( B_\eta^\dagger(r) - r B_\eta^\dagger(1) - \int_0^r \bar{B}_{\eta c, x}(s) \{ \int_0^1 \bar{B}_{\eta c, x}^2(s) \}^{-1} \int \bar{B}_{\eta c, x}(s) dB_\eta^\dagger(s) \right) \Big| B_1 \\ &= \omega_{y|x}^{1/2} F^\dagger(r, c_x) | B_1 \end{aligned}$$

in  $\mathcal{D}$ , where  $F^\dagger(r, c_x) := B_\eta^\dagger(r) - r B_\eta^\dagger(1) - \int_0^r \bar{B}_{\eta c, x}(s) \{ \int_0^1 \bar{B}_{\eta c, x}(s)^2 \}^{-1} \int_0^1 \bar{B}_{\eta c, x}(s) dB_\eta^\dagger(s)$ ,  $r \in [0, 1]$ , and convergence is joint with (S.30). Moreover, using the previous convergence results

we have that,

$$\begin{aligned}
s_y^{*2} &= T^{-1} \sum_{t=1}^T (y_t^* - \bar{y}^*)^2 - T^{-1} \frac{\{T^{-1} \sum_{t=1}^T \hat{x}_{t-1} y_t^*\}^2}{T^{-2} \sum_{t=1}^T \hat{x}_{t-1}^2} + o_p^*(1) \\
&= T^{-1} \sum_{t=1}^T y_t^{*2} + o_p^*(1) = T^{-1} \sum_{t=1}^T w_t^2 \hat{e}_t^2 + o_p^*(1) \\
&= T^{-1} \sum_{t=1}^T \hat{e}_t^2 + T^{-1} \sum_{t=1}^T (w_t^2 - 1) \hat{e}_t^2 + o_p^*(1) \\
&= T^{-1} \sum_{t=1}^T \hat{e}_t^2 + o_p^*(1)
\end{aligned}$$

because  $E^* \{T^{-1} \sum_{t=1}^T (w_t^2 - 1) \hat{e}_t^2\}^2 = 2T^{-2} \sum_{t=1}^T \hat{e}_t^4 = o_p(1)$  under the assumption that the fourth moments are finite. We conclude that  $s_y^{*2} \xrightarrow{w^*} h_{32}^2 f_2 + h_{33}^2 f_3 = \omega_{y|x}$  and, by the CMT, that  $S^* \xrightarrow{w^*} \int F^\dagger(r, c_x)^2 dr | B_1$  jointly with (S.30). Finally,  $E(g(\int F^\dagger(r, c_x)^2 dr) | B_1)$  and  $E(g(\int F(r, c_x)^2 dr) | B_1)$  are a.s. equal to the the same measurable function of  $B_1$ , for every fixed continuous real function  $g$ , because  $(F^\dagger, B_1)$  and  $(F, B_1)$  have the same distribution. This allows us to replace  $\int F^\dagger(r, c_x)^2 dr$  by  $\int F(r, c_x)^2 dr$  in the limit of  $S^*$ . ■

**Proof of Corollary 1:** The asymptotic validity of the bootstrap rests on the result that, as  $T \rightarrow \infty$ ,  $S$  conditional on  $x$ , under  $H_u/H_x$ , and  $S^*$  conditional on the data, under all considered hypotheses, jointly converge weakly to the same random measure.

By Theorem 4, it holds that  $[E_x f(S), E^* f(S^*)]' \xrightarrow{w} [E\{f(S_\infty) | B_1\}, E\{f(S_\infty) | B_1\}]'$  under  $H_u/H_x$ , for all continuous bounded real functions  $f$ , where  $S_\infty := \int F(r, c_x)^2 dr$ . This implies weak convergence of the (random) cumulative distribution functions (or processes) of  $S$  given  $x$  and  $S^*$  given the data, see e.g. Daley and Vere-Jones (2008, pp.143-144). Specifically, if  $G$  denotes the cumulative process of  $S_\infty$  conditional on  $B_1$  (i.e.,  $G(z) := P(S_\infty \leq z | B_1)$ , all  $z$ ), then  $[P_x(S \leq \cdot), P^*(S^* \leq \cdot)]' \xrightarrow{w} [G, G]'$  in  $\mathcal{D}(\mathbb{R}) \times \mathcal{D}(\mathbb{R})$ . As the distribution of  $S_\infty$  conditional on  $B_1$  is atomless a.s. (this follows from the representation of the distribution in question as the distribution of an infinite weighted sum of  $\chi^2$  variables, similarly to Nyblom, 1989, and Rao and Swift, 2006, pp.472-473) and so  $G$  is sample-path continuous a.s., the latter convergence holds also in  $\mathcal{D}^2(\mathbb{R})$  and implies that  $\sup_{z \in \mathbb{R}} |P_x(S \leq z) - P^*(S^* \leq z)| = o_p(1)$ . Therefore, if  $G_T$  denotes the cumulative process of  $S$  conditional on  $x$  (i.e.,  $G_T(z) := P_x(S \leq z)$ , all  $z$ ), then  $P^*(S^* \leq S) = G_T(S) + o_p(1)$ ;

here we have used the fact that  $P^*(S^* \leq z)|_{z=S} = P^*(S^* \leq S)$  due to the measurability of  $S$  with respect to the data.

Further, define the quantile transformation using the right-continuous version of the generalised inverse. Then  $\{G_T(S) \leq \theta\} = \{S \leq G_T^{-1}(\theta)\}$  for all  $\theta \in (0, 1)$ . As the quantile transformation is continuous in the Skorokhod metric, it holds that  $(G_T, G_T^{-1}) \xrightarrow{w} (G, G^{-1})$  in  $\mathcal{D}^2(\mathbb{R})$ . Therefore, for every  $\theta \in (0, 1)$  where  $G^{-1}$  is a.s. continuous,  $(G_T, G_T^{-1}(\theta)) \xrightarrow{w} (G, G^{-1}(\theta))$  in  $\mathcal{D}^2(\mathbb{R}) \times \mathbb{R}$  and

$$P_x(G_T(S) \leq \theta) = P_x(S \leq G_T^{-1}(\theta)) = G_T(G_T^{-1}(\theta)) \xrightarrow{w} G(G^{-1}(\theta)) = \theta$$

a.s., the second equality by the  $\sigma(x)$ -measurability of  $G_T^{-1}(\theta)$ , and the same convergence holds in probability as the limit is a constant. Since such  $\theta$  are all but countably many, we can conclude that  $G_T(S)|x \xrightarrow{w_p} U[0, 1]$ , and since  $P^*(S^* \leq S) = G_T(S) + o_p(1)$ , by (S.29) also  $P^*(S^* \leq S)|x \xrightarrow{w_p} U[0, 1]$ . Finally, by the bounded convergence theorem, the unconditional convergence  $P^*(S^* \leq S) \xrightarrow{w} U[0, 1]$  follows. The statements in the corollary can now be obtained by taking  $1 - P^*(S^* \leq S)$ .

## References

- Billingsley P. (1999). *Convergence of Probability Measures*. John Wiley & Sons, New York.
- Brown, B. (1971). Martingale central limit theorems. *The Annals of Mathematical Statistics* 42, 59–66.
- Crimaldi, I. and L. Pratelli (2005). Convergence results for conditional expectations. *Bernoulli* 11, 737–745.
- Daley, D. and D. Vere-Jones (2008). *An Introduction to the Theory of Point Processes, Volume II: General Theory and Structure*, Springer-Verlag, New York.
- Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association* 84, 223-230.



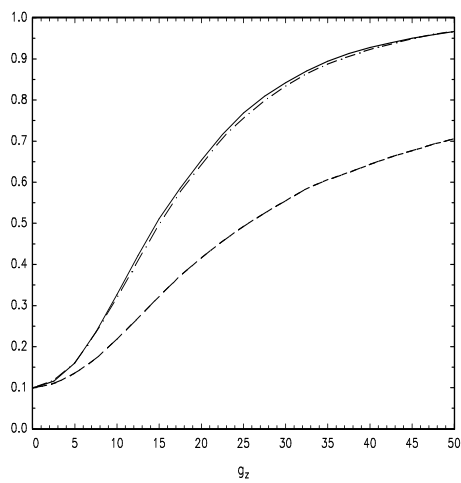
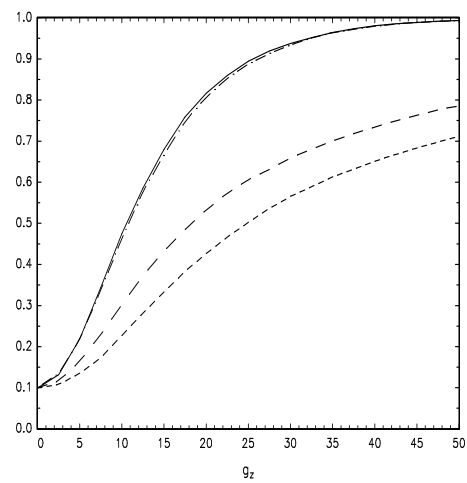
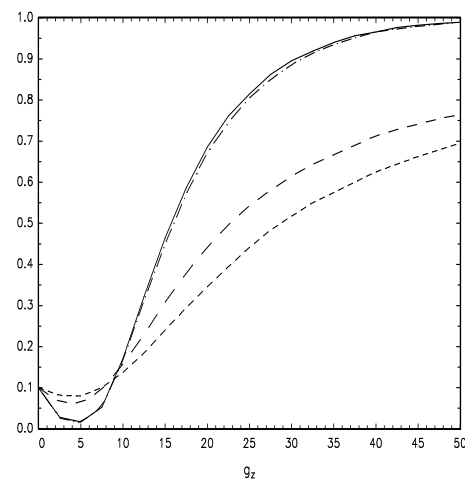
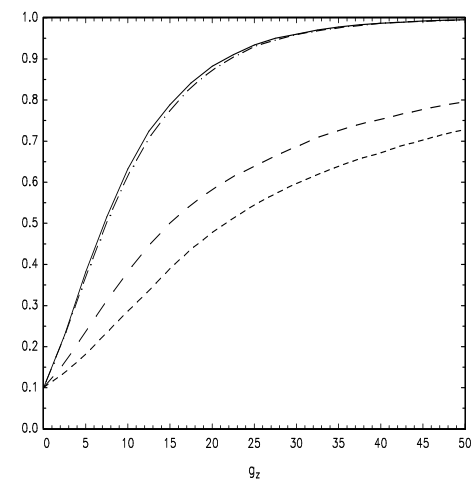
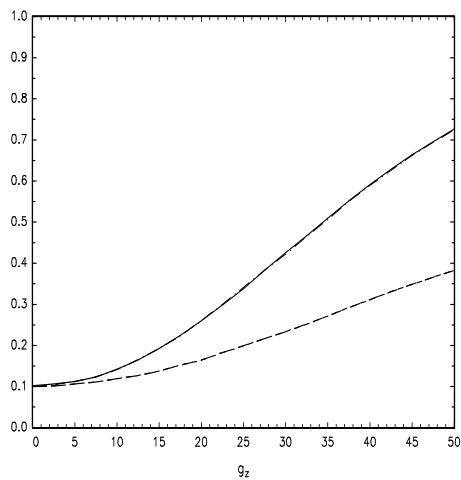
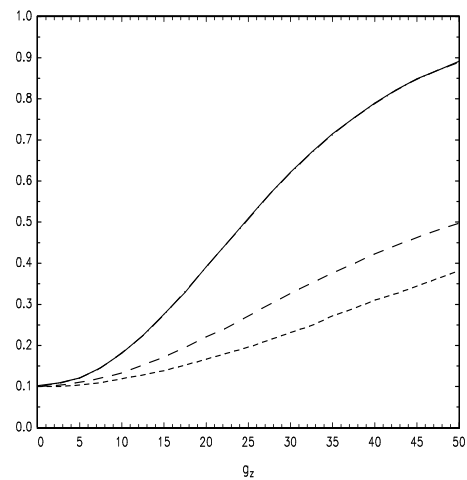
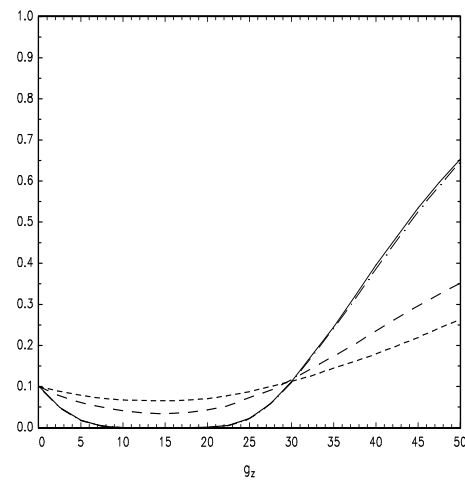
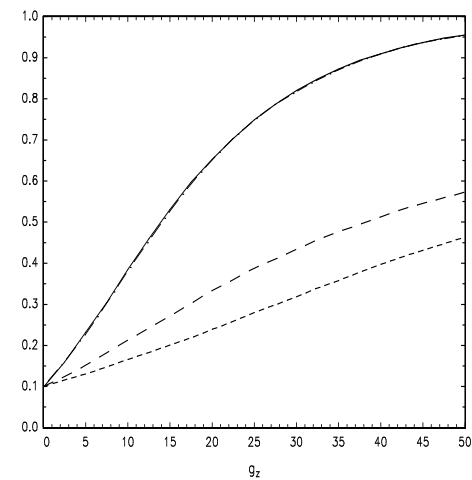
Rao, M.M. and R. J. Swift (2006). Probability Theory with Applications. New York: Springer.

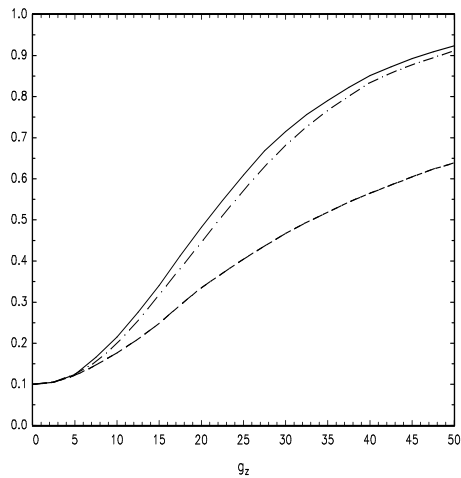
Table S1. Finite sample rejection frequencies of  $S_B$  (power) and  $IV_{comb}$  (size) under volatility shifts:  
 $T = 200$ ,  $g_x = 0$ ,  $g_z = 25$ ,  $d_{it} = 1(t \leq \lfloor \tau T \rfloor) + \sigma_i 1(t > \lfloor \tau T \rfloor)$ ,  $i = 1, 2, 3$

$\sigma_1$	$\sigma_2$	$\sigma_3$	$c_x = c_z = 0$				$c_x = c_z = 5$				$c_x = c_z = 10$				
			$\tau = 0.3$		$\tau = 0.7$		$\tau = 0.3$		$\tau = 0.7$		$\tau = 0.3$		$\tau = 0.7$		
			$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	
1	1	1	0.910	0.712	0.910	0.712	0.742	0.381	0.742	0.381	0.568	0.244	0.568	0.244	
		4	0.478	0.444	0.585	0.511	0.252	0.162	0.308	0.202	0.174	0.130	0.198	0.145	
		$\frac{1}{4}$	0.970	0.760	0.944	0.739	0.880	0.487	0.828	0.420	0.754	0.340	0.688	0.277	
	4	4	1	0.997	0.843	0.977	0.763	0.985	0.634	0.919	0.537	0.960	0.478	0.842	0.403
			4	0.905	0.738	0.815	0.624	0.761	0.401	0.612	0.349	0.585	0.241	0.462	0.229
			$\frac{1}{4}$	0.999	0.854	0.987	0.776	0.995	0.670	0.947	0.567	0.986	0.533	0.895	0.437
		$\frac{1}{4}$	1	0.656	0.556	0.864	0.705	0.469	0.252	0.661	0.348	0.340	0.180	0.481	0.219
			4	0.245	0.275	0.534	0.495	0.153	0.127	0.251	0.190	0.131	0.117	0.168	0.140
			$\frac{1}{4}$	0.817	0.638	0.904	0.735	0.641	0.351	0.754	0.389	0.482	0.247	0.588	0.247
	4	1	1	0.907	0.722	0.912	0.685	0.739	0.383	0.745	0.384	0.569	0.240	0.576	0.253
			4	0.464	0.427	0.602	0.386	0.254	0.160	0.317	0.170	0.175	0.127	0.204	0.140
			$\frac{1}{4}$	0.971	0.795	0.942	0.764	0.885	0.552	0.823	0.517	0.751	0.412	0.680	0.375
4		4	1	0.996	0.856	0.968	0.755	0.982	0.643	0.907	0.580	0.956	0.477	0.828	0.461
			4	0.896	0.738	0.781	0.555	0.754	0.386	0.577	0.324	0.579	0.229	0.432	0.226
			$\frac{1}{4}$	0.999	0.870	0.978	0.782	0.993	0.691	0.940	0.637	0.983	0.548	0.882	0.536
		$\frac{1}{4}$	1	0.679	0.551	0.886	0.662	0.487	0.239	0.688	0.322	0.351	0.167	0.505	0.203
			4	0.253	0.260	0.576	0.361	0.158	0.127	0.279	0.154	0.135	0.118	0.178	0.132
			$\frac{1}{4}$	0.826	0.685	0.919	0.751	0.660	0.400	0.771	0.460	0.494	0.287	0.601	0.307
$\frac{1}{4}$		1	1	0.909	0.695	0.914	0.719	0.744	0.377	0.733	0.367	0.573	0.257	0.567	0.234
			4	0.494	0.504	0.584	0.569	0.255	0.201	0.291	0.257	0.174	0.144	0.190	0.177
			$\frac{1}{4}$	0.975	0.721	0.943	0.733	0.874	0.421	0.828	0.385	0.755	0.288	0.687	0.246
	4	4	1	0.996	0.835	0.979	0.760	0.988	0.621	0.913	0.498	0.965	0.475	0.838	0.348
			4	0.920	0.765	0.824	0.662	0.791	0.471	0.614	0.387	0.606	0.305	0.466	0.256
			$\frac{1}{4}$	0.999	0.842	0.989	0.767	0.996	0.637	0.946	0.507	0.988	0.493	0.894	0.359
		$\frac{1}{4}$	1	0.603	0.571	0.855	0.719	0.444	0.289	0.649	0.357	0.323	0.215	0.468	0.224
			4	0.214	0.339	0.515	0.553	0.150	0.156	0.235	0.248	0.129	0.133	0.161	0.170
			$\frac{1}{4}$	0.785	0.608	0.897	0.736	0.596	0.322	0.750	0.368	0.448	0.240	0.585	0.235

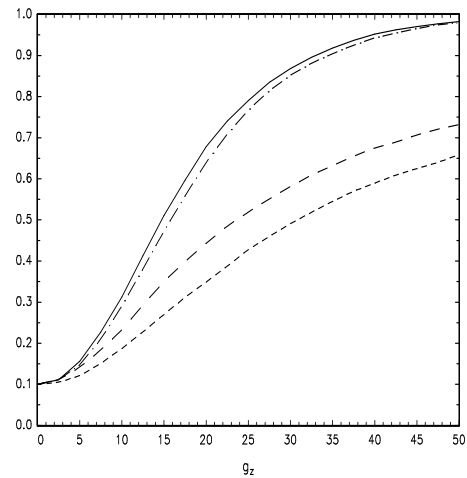
Table S2. Finite sample rejection frequencies of  $S_B$  (power) and  $IV_{comb}$  (size) under volatility shifts:  
 $T = 200, g_x = 0, g_z = 50, d_{it} = 1(t \leq \lfloor \tau T \rfloor) + \sigma_i 1(t > \lfloor \tau T \rfloor), i = 1, 2, 3$

$\sigma_1$	$\sigma_2$	$\sigma_3$	$c_x = c_z = 0$				$c_x = c_z = 5$				$c_x = c_z = 10$				
			$\tau = 0.3$		$\tau = 0.7$		$\tau = 0.3$		$\tau = 0.7$		$\tau = 0.3$		$\tau = 0.7$		
			$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	$S_B$	$IV_{comb}$	
1	1	1	0.987	0.804	0.987	0.804	0.944	0.545	0.944	0.545	0.866	0.393	0.866	0.393	
		4	0.761	0.630	0.848	0.682	0.527	0.276	0.607	0.345	0.356	0.175	0.414	0.221	
		$\frac{1}{4}$	0.996	0.830	0.992	0.815	0.981	0.617	0.968	0.576	0.949	0.480	0.924	0.425	
	4	4	1	1.000	0.860	0.996	0.813	0.997	0.686	0.976	0.634	0.992	0.555	0.944	0.518
			4	0.984	0.821	0.956	0.749	0.946	0.566	0.857	0.517	0.871	0.390	0.740	0.374
			$\frac{1}{4}$	1.000	0.866	0.998	0.817	0.999	0.700	0.983	0.645	0.996	0.572	0.958	0.532
		$\frac{1}{4}$	1	0.886	0.714	0.973	0.804	0.777	0.421	0.908	0.522	0.639	0.298	0.809	0.365
			4	0.465	0.458	0.796	0.678	0.293	0.185	0.530	0.315	0.210	0.142	0.339	0.205
			$\frac{1}{4}$	0.951	0.767	0.983	0.818	0.882	0.526	0.948	0.558	0.793	0.404	0.882	0.400
	4	1	1	0.988	0.813	0.987	0.785	0.942	0.548	0.946	0.546	0.860	0.387	0.874	0.398
			4	0.759	0.614	0.861	0.577	0.520	0.258	0.625	0.271	0.353	0.161	0.431	0.181
			$\frac{1}{4}$	0.996	0.845	0.992	0.820	0.981	0.651	0.970	0.633	0.948	0.524	0.925	0.508
4		4	1	0.999	0.873	0.994	0.791	0.996	0.690	0.977	0.647	0.990	0.542	0.944	0.545
			4	0.982	0.828	0.943	0.700	0.940	0.554	0.845	0.490	0.866	0.365	0.726	0.364
			$\frac{1}{4}$	1.000	0.877	0.996	0.800	0.998	0.704	0.984	0.666	0.994	0.566	0.961	0.571
		$\frac{1}{4}$	1	0.891	0.713	0.978	0.774	0.784	0.404	0.922	0.495	0.649	0.270	0.826	0.334
			4	0.487	0.437	0.829	0.552	0.297	0.167	0.575	0.225	0.218	0.135	0.370	0.156
			$\frac{1}{4}$	0.957	0.797	0.985	0.822	0.894	0.572	0.952	0.612	0.797	0.451	0.889	0.467
$\frac{1}{4}$		1	1	0.987	0.792	0.987	0.808	0.947	0.549	0.941	0.538	0.869	0.409	0.857	0.381
			4	0.773	0.677	0.847	0.729	0.536	0.353	0.589	0.436	0.356	0.241	0.400	0.293
			$\frac{1}{4}$	0.997	0.799	0.992	0.815	0.980	0.575	0.969	0.552	0.949	0.440	0.924	0.392
	4	4	1	0.999	0.861	0.997	0.815	0.998	0.691	0.975	0.619	0.993	0.574	0.939	0.486
			4	0.985	0.832	0.958	0.773	0.955	0.617	0.852	0.554	0.884	0.462	0.736	0.413
			$\frac{1}{4}$	1.000	0.861	0.998	0.820	0.999	0.697	0.983	0.621	0.997	0.583	0.954	0.491
		$\frac{1}{4}$	1	0.849	0.716	0.970	0.813	0.749	0.464	0.901	0.532	0.623	0.355	0.798	0.368
			4	0.407	0.540	0.790	0.730	0.285	0.261	0.498	0.419	0.204	0.200	0.320	0.281
			$\frac{1}{4}$	0.946	0.735	0.982	0.821	0.851	0.496	0.944	0.542	0.761	0.389	0.878	0.379

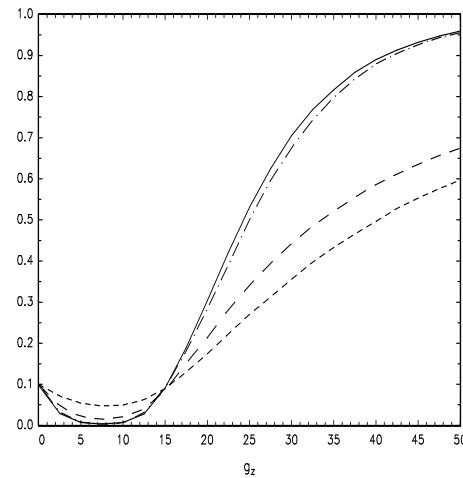
(a)  $c = 5, \sigma_{xy} = 0, \sigma_{zy} = 0$ (b)  $c = 5, \sigma_{xy} = -0.7, \sigma_{zy} = 0$ (c)  $c = 5, \sigma_{xy} = -0.7, \sigma_{zy} = -0.7$ (d)  $c = 5, \sigma_{xy} = -0.7, \sigma_{zy} = 0.7$ (e)  $c = 20, \sigma_{xy} = 0, \sigma_{zy} = 0$ (f)  $c = 20, \sigma_{xy} = -0.7, \sigma_{zy} = 0$ (g)  $c = 20, \sigma_{xy} = -0.7, \sigma_{zy} = -0.7$ (h)  $c = 20, \sigma_{xy} = -0.7, \sigma_{zy} = 0.7$ Figure S1. Asymptotic rejection frequencies of  $S$ ,  $S_B$  (power) and  $t_u$ ,  $Q$  (size):  $g_x = 0, c_x = c_z = c$ ;
$$S: \cdots, S_B: \text{—}, t_u: \text{---}, Q: \text{--}$$



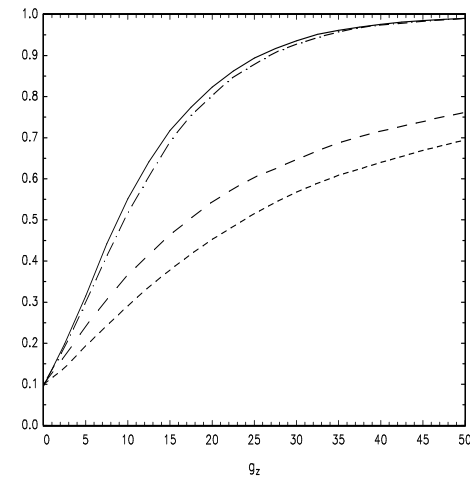
(a)  $c_x = 0, c_z = 10,$   
 $\sigma_{xy} = 0, \sigma_{zy} = 0$



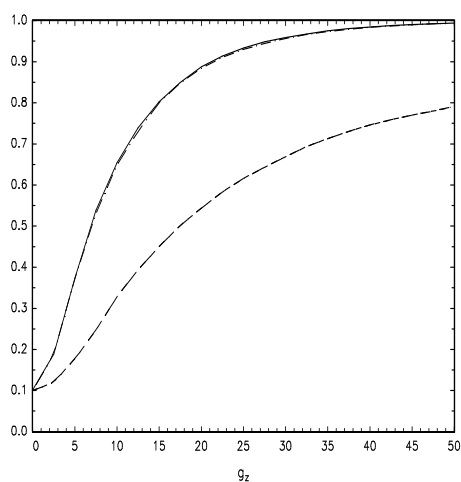
(b)  $c_x = 0, c_z = 10,$   
 $\sigma_{xy} = -0.7, \sigma_{zy} = 0$



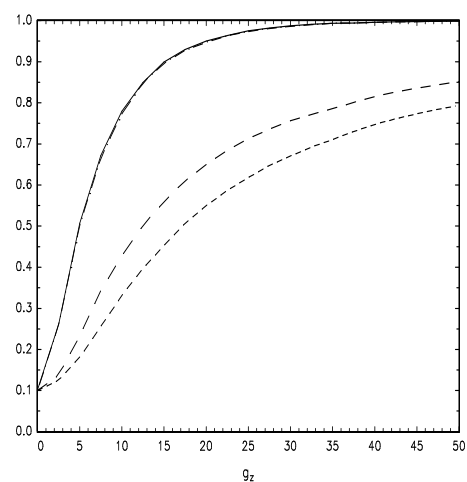
(c)  $c_x = 0, c_z = 10,$   
 $\sigma_{xy} = -0.7, \sigma_{zy} = -0.7$



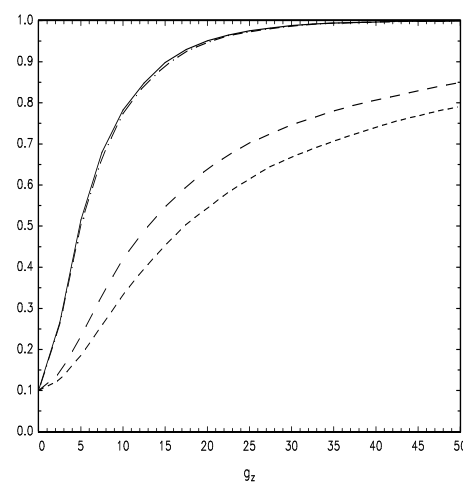
(d)  $c_x = 0, c_z = 10,$   
 $\sigma_{xy} = -0.7, \sigma_{zy} = 0.7$



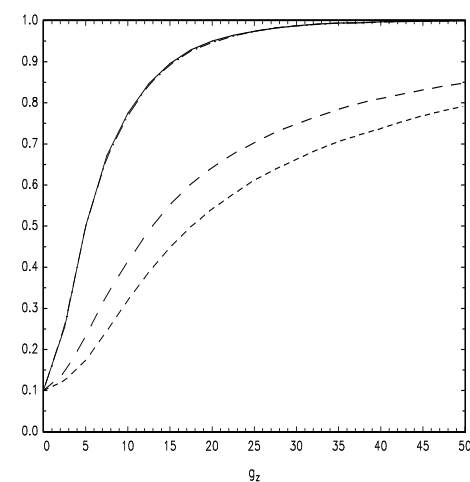
(e)  $c_x = 10, c_z = 0,$   
 $\sigma_{xy} = 0, \sigma_{zy} = 0$



(f)  $c_x = 10, c_z = 0,$   
 $\sigma_{xy} = -0.7, \sigma_{zy} = 0$

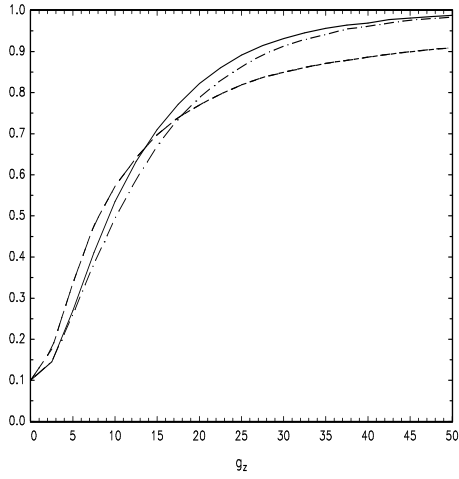


(g)  $c_x = 10, c_z = 0,$   
 $\sigma_{xy} = -0.7, \sigma_{zy} = -0.7$

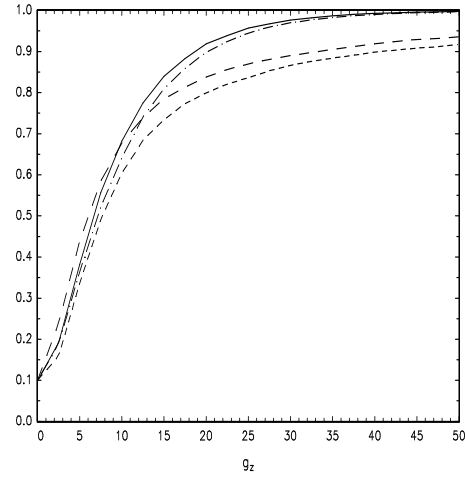


(h)  $c_x = 10, c_z = 0,$   
 $\sigma_{xy} = -0.7, \sigma_{zy} = 0.7$

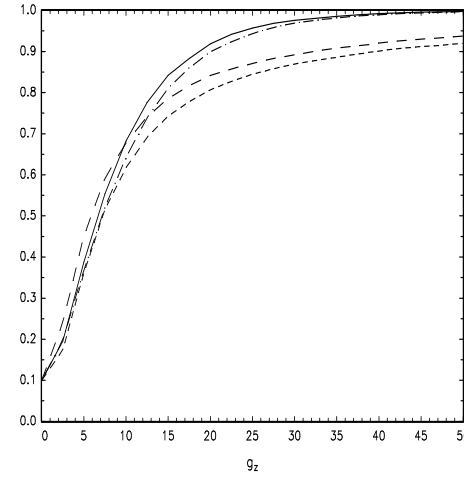
Figure S2. Asymptotic rejection frequencies of  $S$ ,  $S_B$  (power) and  $t_u$ ,  $Q$  (size):  $g_x = 0$ ;  
 $S$ :  $-\cdot-$ ,  $S_B$ :  $—$ ,  $t_u$ :  $---$ ,  $Q$ :  $---$



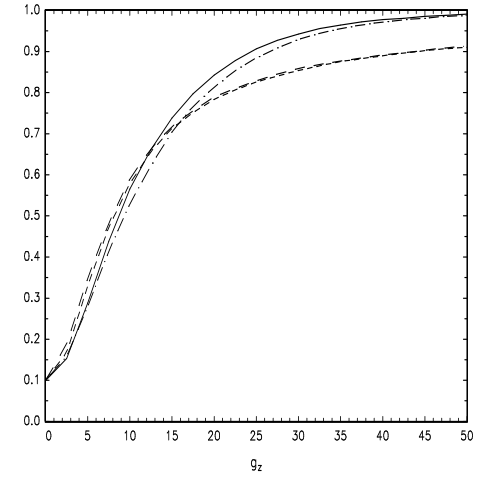
(a)  $\sigma_{xz} = 0.5,$   
 $\sigma_{xy} = 0, \sigma_{zy} = 0$



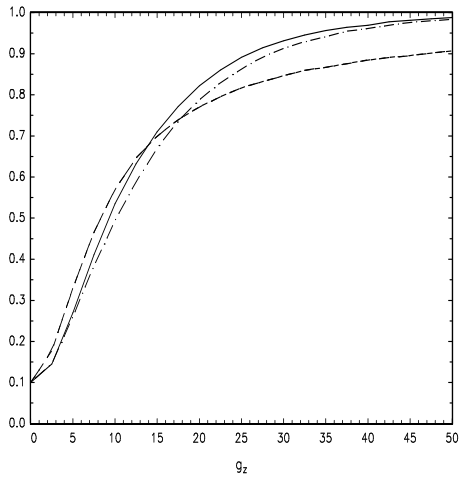
(b)  $\sigma_{xz} = 0.5,$   
 $\sigma_{xy} = -0.7, \sigma_{zy} = 0$



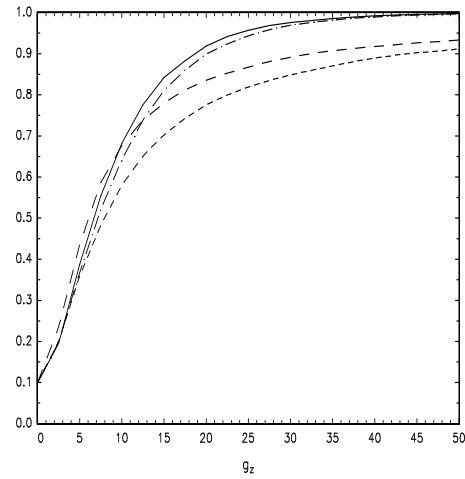
(c)  $\sigma_{xz} = 0.5,$   
 $\sigma_{xy} = -0.7, \sigma_{zy} = -0.7$



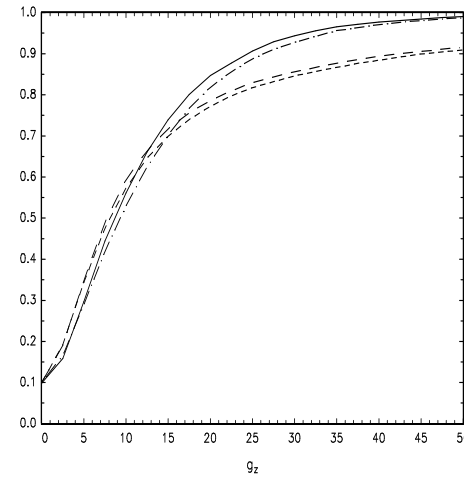
(d)  $\sigma_{xz} = 0.5,$   
 $\sigma_{xy} = -0.35, \sigma_{zy} = 0.35$



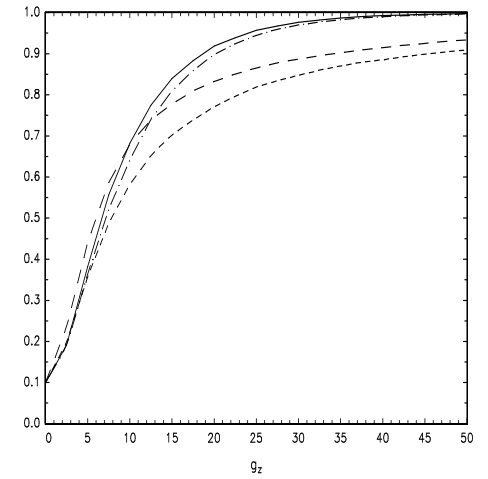
(e)  $\sigma_{xz} = -0.5,$   
 $\sigma_{xy} = 0, \sigma_{zy} = 0$



(f)  $\sigma_{xz} = -0.5,$   
 $\sigma_{xy} = -0.7, \sigma_{zy} = 0$

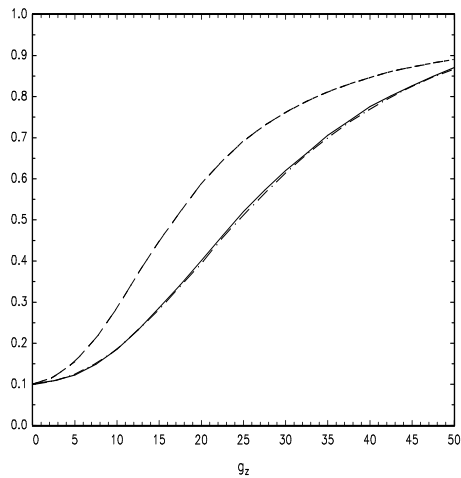


(g)  $\sigma_{xz} = -0.5,$   
 $\sigma_{xy} = -0.35, \sigma_{zy} = -0.35$

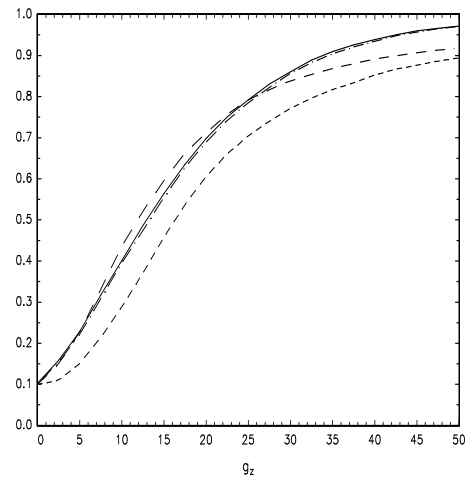


(h)  $\sigma_{xz} = -0.5,$   
 $\sigma_{xy} = -0.7, \sigma_{zy} = 0.7$

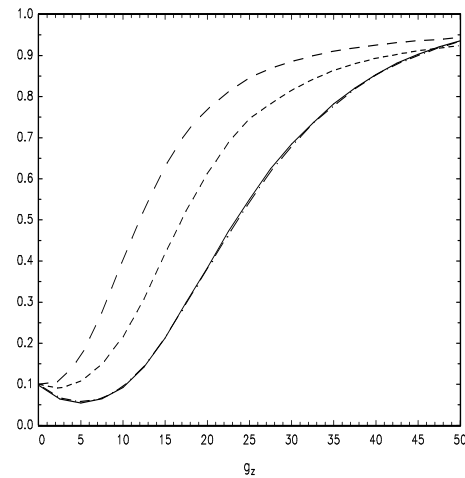
Figure S3. Asymptotic rejection frequencies of  $S$ ,  $S_B$  (power) and  $t_u$ ,  $Q$  (size):  $g_x = 0, c_x = c_z = 0$ ;  
 $S$ :  $-\cdot-$ ,  $S_B$ :  $—$ ,  $t_u$ :  $---$ ,  $Q$ :  $---$



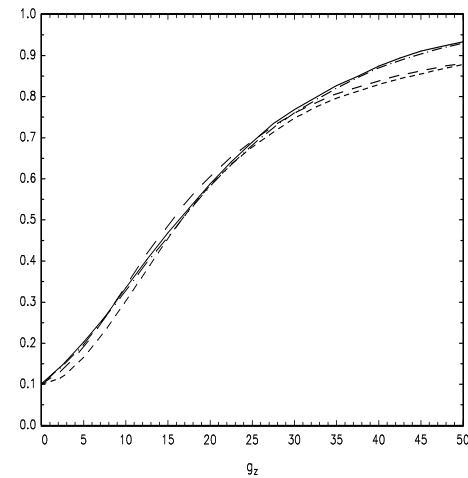
(a)  $\sigma_{xz} = 0.5$ ,  
 $\sigma_{xy} = 0, \sigma_{zy} = 0$



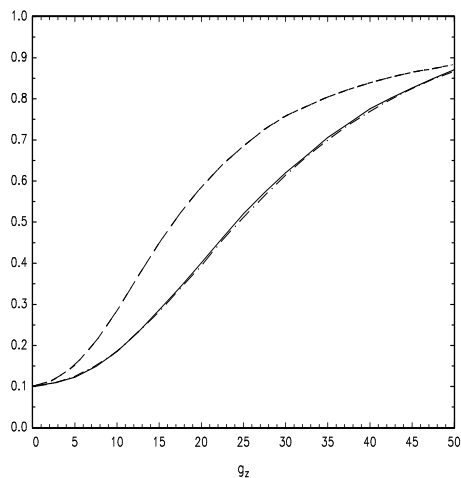
(b)  $\sigma_{xz} = 0.5$ ,  
 $\sigma_{xy} = -0.7, \sigma_{zy} = 0$



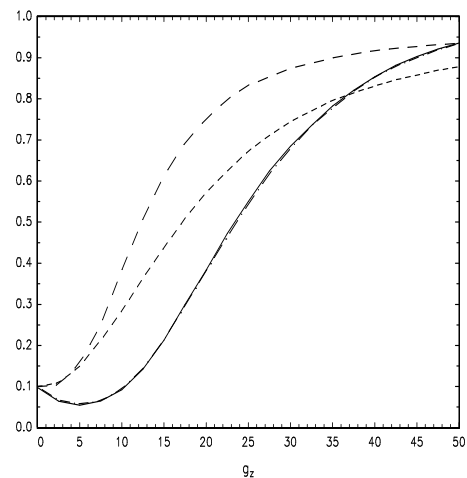
(c)  $\sigma_{xz} = 0.5$ ,  
 $\sigma_{xy} = -0.7, \sigma_{zy} = -0.7$



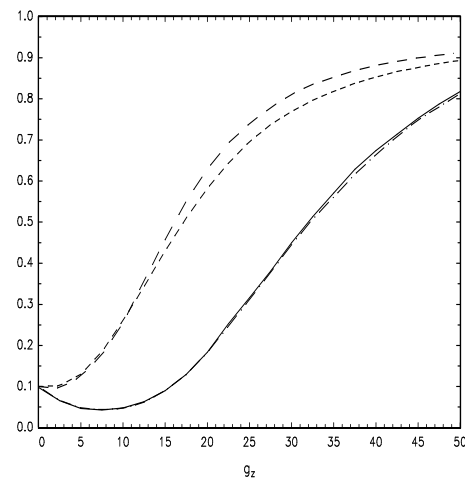
(d)  $\sigma_{xz} = 0.5$ ,  
 $\sigma_{xy} = -0.35, \sigma_{zy} = 0.35$



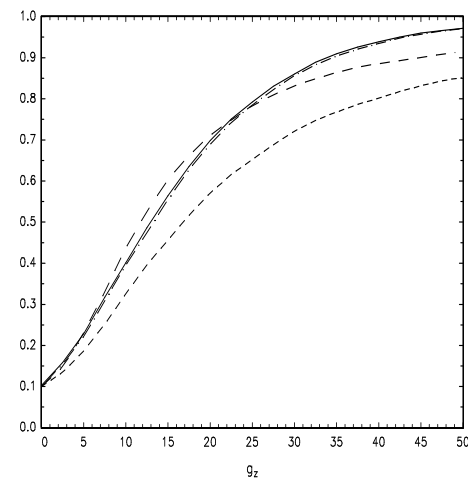
(e)  $\sigma_{xz} = -0.5$ ,  
 $\sigma_{xy} = 0, \sigma_{zy} = 0$



(f)  $\sigma_{xz} = -0.5$ ,  
 $\sigma_{xy} = -0.7, \sigma_{zy} = 0$

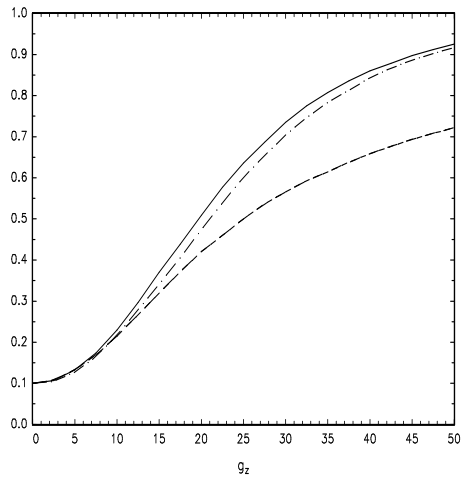


(g)  $\sigma_{xz} = -0.5$ ,  
 $\sigma_{xy} = -0.35, \sigma_{zy} = -0.35$

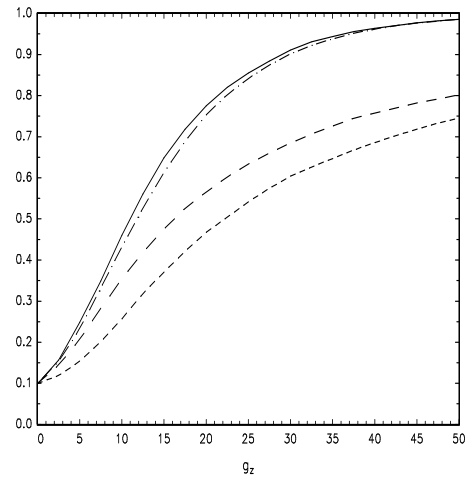


(h)  $\sigma_{xz} = -0.5$ ,  
 $\sigma_{xy} = -0.7, \sigma_{zy} = 0.7$

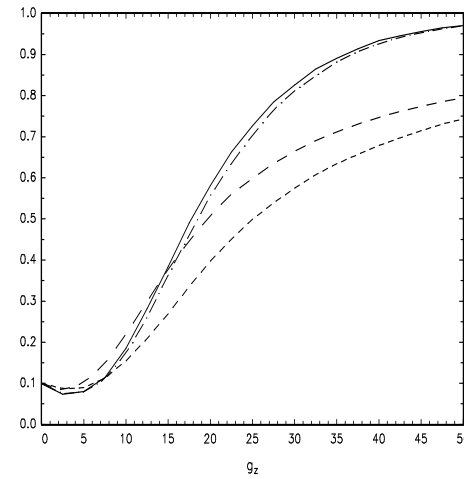
Figure S4. Asymptotic rejection frequencies of  $S$ ,  $S_B$  (power) and  $t_u$ ,  $Q$  (size):  $g_x = 0$ ,  $c_x = c_z = 10$ ;  
 $S$ :  $-\cdot-$ ,  $S_B$ :  $—$ ,  $t_u$ :  $----$ ,  $Q$ :  $----$



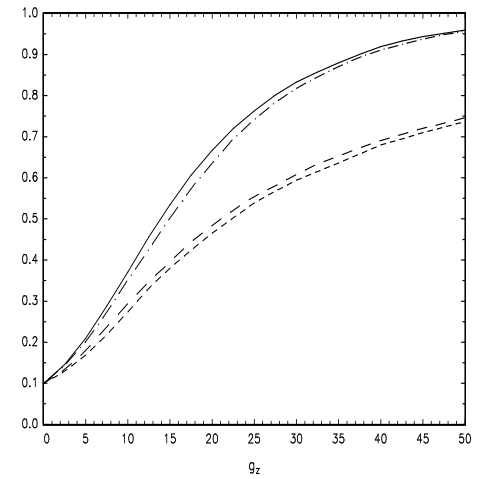
(a)  $\sigma_{xz} = 0.5$ ,  
 $\sigma_{xy} = 0, \sigma_{zy} = 0$



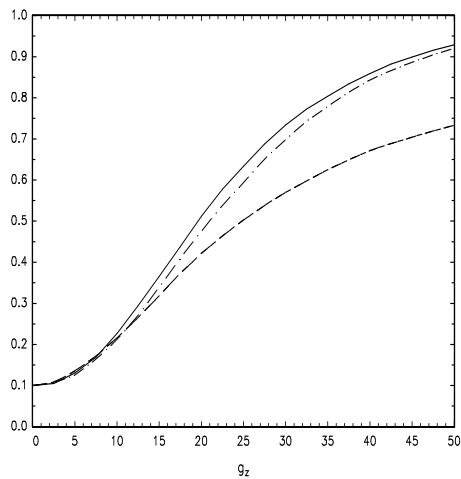
(b)  $\sigma_{xz} = 0.5$ ,  
 $\sigma_{xy} = -0.7, \sigma_{zy} = 0$



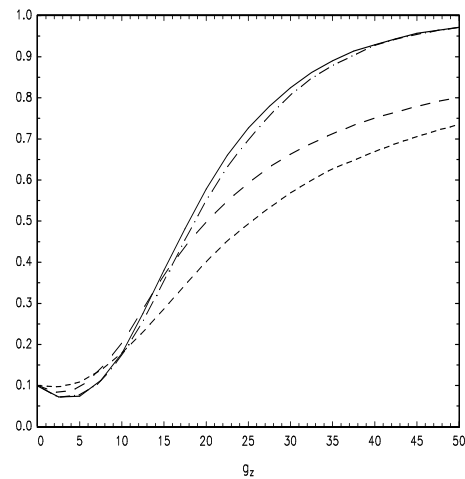
(c)  $\sigma_{xz} = 0.5$ ,  
 $\sigma_{xy} = -0.7, \sigma_{zy} = -0.7$



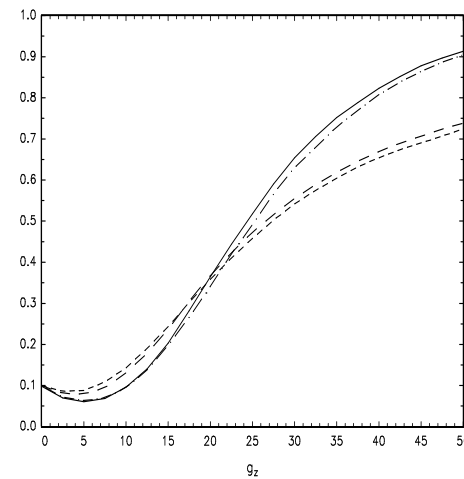
(d)  $\sigma_{xz} = 0.5$ ,  
 $\sigma_{xy} = -0.35, \sigma_{zy} = 0.35$



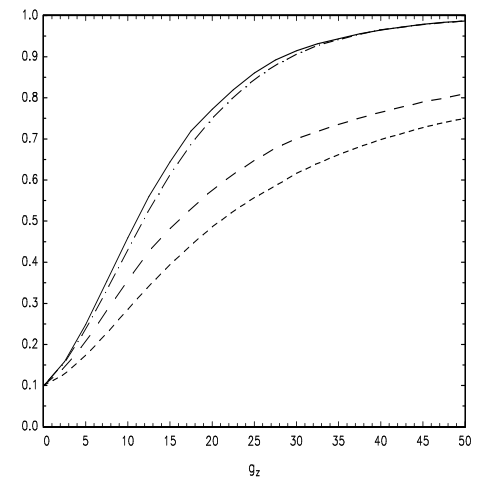
(e)  $\sigma_{xz} = -0.5$ ,  
 $\sigma_{xy} = 0, \sigma_{zy} = 0$



(f)  $\sigma_{xz} = -0.5$ ,  
 $\sigma_{xy} = -0.7, \sigma_{zy} = 0$



(g)  $\sigma_{xz} = -0.5$ ,  
 $\sigma_{xy} = -0.35, \sigma_{zy} = -0.35$



(h)  $\sigma_{xz} = -0.5$ ,  
 $\sigma_{xy} = -0.7, \sigma_{zy} = 0.7$

Figure S5. Asymptotic rejection frequencies of  $S$ ,  $S_B$  (power) and  $t_u$ ,  $Q$  (size):  $g_x = 0, c_x = 0, c_z = 10$ ;  
 $S$ :  $-\cdot-$ ,  $S_B$ :  $—$ ,  $t_u$ :  $----$ ,  $Q$ :  $---$



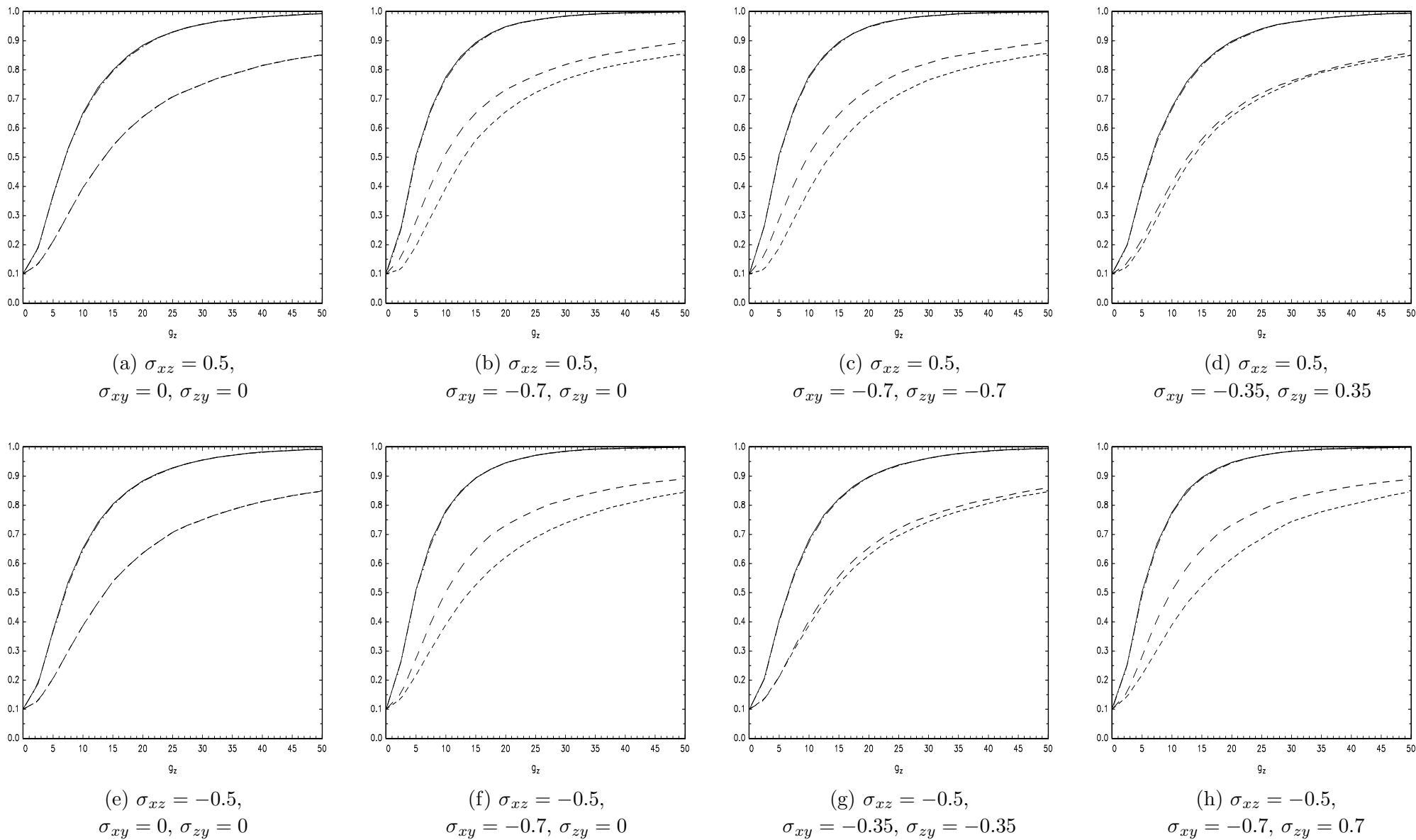


Figure S6. Asymptotic rejection frequencies of  $S$ ,  $S_B$  (power) and  $t_u$ ,  $Q$  (size):  $g_x = 0$ ,  $c_x = 10$ ,  $c_z = 0$ ;  
 $S$ :  $-\cdot-$ ,  $S_B$ :  $—$ ,  $t_u$ :  $----$ ,  $Q$ :  $---$