A Bootstrap Stationarity Test for Predictive Regression Invalidity*

Iliyan Georgiev[a], David I. Harvey[b], Stephen J. Leybourne[b] and A.M. Robert Taylor[c]

[a]Department of Statistics, University of Bologna

[b]School of Economics, University of Nottingham

[c]Essex Business School, University of Essex

November 30, 2015

Abstract

We examine how the familiar spurious regression problem can manifest itself in the context of recently proposed predictability tests. For these tests to provide asymptotically valid inference, account has to be taken of the degree of persistence of the putative predictors. Failure to do so can lead to spurious over-rejections of the no predictability null hypothesis. A number of methods have been developed to achieve this. However, these approaches all make an underlying assumption that any predictability in the variable of interest is purely attributable to the predictors under test, rather than to any unobserved persistent latent variables, themselves uncorrelated with the predictors being tested. We show that where this assumption is violated, something that could very plausibly happen in practice, sizeable (spurious) rejections of the null can occur in cases where the variables under test are not valid predictors. In response, we propose a screening test for predictive regression invalidity based on a stationarity testing approach. In order to allow for an unknown degree of persistence in the putative predictors, and for both conditional and unconditional heteroskedasticity in the data, we implement our proposed test using a fixed regressor wild bootstrap procedure. We establish the asymptotic validity of this bootstrap test, which entails establishing a *conditional* invariance principle along with its bootstrap counterpart, both of which appear to be new to the literature and are likely to have important applications beyond the present context. We also show how our bootstrap test can be used, in conjunction with extant predictability tests, to deliver a two-step feasible procedure. Monte Carlo simulations suggest that our proposed bootstrap methods work well in finite samples. An illustration employing U.S. stock returns data demonstrates the practical usefulness of our procedures.

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

JEL Classification: C12, C32.

*Correspondence to: Robert Taylor, Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. Email: rtaylor@essex.ac.uk.

1 Introduction

Predictive regressions are a widely used tool in applied time series finance and economics, forming the basis for testing Granger causality where one or more variables can be said to cause another. A very common application of predictive regression is in the context of testing the linear rational expectations hypothesis. A core example of this is testing whether future stock returns can be predicted 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. the dividend yield) exhibits persistence behaviour akin to that of a unit root or near unit root autoregressive process, whilst the variable being predicted (e.g. the stock return) resembles a white noise or near white noise process.

In its simplest 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 being tested is that y_t is white noise and, hence, unpredictable; the alternative is that it is not white noise and is in fact predictable from x_{t-1}. Cavanagh *et al.* (1995) show that when the innovation driving the x_t process is correlated with y_t (as is often thought to be case in practice; for example, the stock price is a component of both the return and the dividend yield), then standard t-ratios from a predictive OLS regression of y_t on x_{t-1} which apply critical values appropriate for the unit root case for x_t, will be badly over-sized if in fact x_t is a local to unit root process. This over-size issue can clearly 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 actuality y_t is white noise. Attempting to address this issue, Cavanagh *et al.* (1995) discuss Bonferroni bound-based procedures that yield conservative tests, while Campbell and Yogo (2006) consider a point optimal variant of the t-test and employ confidence belts. Recently, Breitung and Demetrescu (2015) consider variable addition and instrumental variable (IV) methods to correct test size. Near-optimal predictive regression tests can also be found in Elliott *et al.* (2015) and Jansson and Moreira (2006).

Spurious predictability can also arise when there is predictability in y_t, not caused by x_{t-1}, but instead by an unobserved persistent latent variable, z_{t-1}, which itself is uncorrelated with x_{t-1}. In such cases, regression of y_t on the observed, but invalid, potential predictor variable x_{t-1} can lead to serious upward size distortions in the standard predictive regression tests, with the result that the predictability of y_t by the unobserved z_{t-1} is spuriously assigned to x_{t-1}. This issue was considered by Ferson *et al.* (2003a,b) and exemplified by means of Monte Carlo simulation, while Deng (2014) provides a supporting asymptotic analysis of the problem. This fundamental mis-specification problem is also pertinent to the procedures employed by Cavanagh *et al.* (1995), Campbell and Yogo (2006) and Breitung and Demetrescu (2015). In this paper we demonstrate theoretically the potential for spurious predictive regression to arise in the context of a model where x_t and z_t follow similar but uncorrelated persistent processes, which we model as local-to-unity autoregressions, while modelling the coefficient on z_{t-1} as being local-to-zero. We find that spurious rejections in favour of y_t being predicted by x_{t-1} can occur frequently. As a consequence, it is important to be able to identify, *a priori*, whether or not the predictive regression of y_t on x_{t-1} is mis-specified due to the omission of the relevant predictor z_{t-1}.

The approach we adopt in this paper 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 predictive regression, is eliminated from the residuals, and any persistence they display thereafter must be attributable to the unobserved latent variable z_{t-1}, and this would signal invalidity of a predictive regression that employs x_{t-1}. The test for predictive regression invalidity that we propose is based on adapting the co-integration tests of Shin (1994) and Leybourne and McCabe (1994), which are themselves variants of the stationarity test of Kwiatkowski *et al.* (1992) test (KPSS).

An issue that arises with our proposed KPSS-type test is that under the null hypothesis for this test, which is that z_{t-1} plays no role in the data generating process [DGP] for y_t, it has a limit distribution that depends on the local-to-unity parameter in the process for x_t, even though the residuals used in its construction are invariant to the slope coefficient on x_{t-1} in the putative predictive regression. In principle, this makes it very difficult to control the size of the test. However, we will show how a bootstrap procedure which treats x_{t-1} as a fixed regressor (i.e. the observed x_{t-1} is used in calculating the bootstrap analogues of the KPSS-type statistic) can be implemented to yield an asymptotically size-controlled testing strategy. This *fixed regressor bootstrap* type approach is not new to the literature and has been successfully employed in the context of other testing applications by, among others, Gonçalves and Kilian (2004, 2007) and Hansen (2000). Since many financial and economic time series are thought to display non-stationary volatility and/or conditional heteroskedasticity in their driving innovations, it is also important for our proposed testing procedure to be (asymptotically) robust to these effects. In order to achieve this we use a heteroskedasticity-robust variant of the fixed regressor bootstrap along the lines proposed in Hansen (2000) which uses a wild bootstrap scheme to generate bootstrap analogues of y_t. We show that our proposed fixed regressor wild bootstrap (or *heteroskedastic fixed regressor bootstrap* in the terminology of Hansen, 2000) testing procedure has non-trivial local asymptotic power against the same local alternatives that induce a finding of spurious predictive regression of y_t by x_{t-1}, and is therefore of value as a screening tool for potential predictive regression invalidity.

Establishing the large sample validity of our proposed bootstrap method is shown, because of the fixed regressor aspect of its construction, to entail the need to establish a conditional joint invariance principle for the original data and the bootstrap data, which is to the best of our knowledge novel in the literature. This result is likely to have wider uses beyond the specific testing problem considered here, in cases where persistent regressors are used in a fixed regressor bootstrap scenario.

The remainder of the paper is laid out as follows. In section 2 we give the basic model under which we operate and set out the various null and alternative hypotheses regarding the potential predictability of y_t by x_{t-1} and z_{t-1}. To aid lucidity, we expound our approach through a single putative predictor variable, x_t, and a single unobserved latent variable, z_t. For the same reason we also initially abstract away from weak dependence in the innovations driving x_t and z_t. Generalisations to allow for multiple putative predictors, multiple latent variables, and weak dependence are conceptually straightforward and are discussed at various points within the text. In section 3 we derive the asymptotic distributions of standard predictive regression statistics under the various hypotheses, and demonstrate the potential spurious predictive regression problem. Section 4 introduces our proposed stationarity test for predictive regression 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 the predictive regression tests. Section 5 presents the results of a set of finite sample Monte Carlo simulations which investigate the size and power of our proposed procedure, as well as the size of feasible predictive regression tests, and also highlights the value of the predictive regression invalidity test by reporting results where the procedure is used a pre-test prior to application of a predictive regression test. An empirical illustration of the proposed methods to monthly U.S stock returns data is presented in Section 7. Some conclusions are offered in section 8. All proofs are contained in a mathematical appendix.

Before proceeding to the main part of the paper, we first introduce some notation. In what follows, '$\lfloor \cdot \rfloor$' is used to denote the integer part of its argument, '$\mathbb{I}(.)$' denotes the indicator function, '$x := y$' ('$x =: y$') indicates that x is defined by y (y is defined by x). The notation '$\xrightarrow{w}$' denotes weak convergence and '$\xrightarrow{p}$' convergence in probability, in each case as the sample size diverges. For any vector, x, $\|x\|$ denotes the usual Euclidean norm, $\|x\| := (x'x)^{1/2}$ Finally, $\mathcal{D}^k := D_k[0,1]$ denotes the space of right continuous with left limit (càdlàg) $\mathbb{R}^k$-valued functions on $[0,1]$, equipped with the Skorokhod topology.

2 The Predictive Regression Model

The basic DGP we consider for observed y_t is

$$y_t = \alpha_y + \beta_x x_{t-1} + \beta_z z_{t-1} + \epsilon_{yt}, \qquad t = 1, ..., T \tag{1}$$

where x_t is an observed process, while z_t is unobserved, with specifications

$$x_t = \alpha_x + s_{x,t}, \qquad z_t = \alpha_z + s_{z,t}, \qquad t = 0, ..., T \tag{2}$$

$$s_{x,t} = \rho_x s_{x,t-1} + \epsilon_{xt}, \qquad s_{z,t} = \rho_z s_{z,t-1} + \epsilon_{zt}, \qquad t = 1, ..., T \tag{3}$$

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 persistent unit root or local to unit root autoregressive processes. We let $s_{x,0}$ and $s_{z,0}$ be $O_p(1)$ variates. Following Cavanagh *et al.* (1995) and in order to examine the asymptotic local power of the test procedures we discuss, we parameterise β_x and β_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.

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

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

(a) H and D_t are the 3×3 non-stochastic matrices

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

such that

$$HD_t = \begin{bmatrix} h_{11}d_{1t} & 0 & 0 \\ 0 & h_{22}d_{2t} & 0 \\ h_{31}d_{1t} & h_{32}d_{2t} & h_{33}d_{3t} \end{bmatrix}$$

with HH' strictly positive definite. The volatility terms d_{it} satisfy $d_{it} = d_i(t/T)$, where $d_i(\cdot) \in \mathcal{D}$ are non-stochastic, strictly positive functions.

(b) e_t is a 3×1 vector martingale difference sequence [m.d.s.] with respect to a filtration $\mathcal{F}_t$, 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 above essentially coincides with Assumption 2 of Boswijk *et al.* (2015), except for the omission of their conditions (b)ii and (b)iii, and the special structure of the matrix HD_t imposed here. Specifically, the structure of H given above imposes zero correlation between ϵ_{xt} and ϵ_{zt}, while allowing ϵ_{yt} to be potentially correlated with ϵ_{xt} and/or ϵ_{zt}. Assumption 1 implies that ϵ_t is a vector martingale difference sequence 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 $\Omega_t := E(\epsilon_t \epsilon_t') = (HD_t)(HD_t)' > 0$.[1] Stationary conditional heteroskedasticity and non-stationary unconditional volatility are obtained as special cases with $d_i(\cdot) = d_i$, $i = 1, 2, 3$, (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.

Remark 2. As discussed in Cavaliere, Rahbek and Taylor (2010), Assumption 1(a) implies that the elements of Ω_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 ϵ_t (subject to the structure imposed by H), including single or multiple variance or covariance shifts, variances which follow a broken trend, and smooth transition variance shifts; see also the discussion following Assumption 1 of Breitung and Demetrescu (2015)

[1] Notice that the assumption that $E(e_t e_t') = I_3$ made in part (b)i is without loss of generality and is made only to simplify notation.

who allow for similar conditions on a version of the model in (1), (2) and (3) that does not include z. Again for a version of the model in (1), (2) and (3) that omits z, Assumption 2 (b) coincides with the martingale difference conditions stated in Assumption 1 of Breitung and Demetrescu (2015), except that the cross product moment summability condition given there is not required in the context of (1), (2) and (3) because we do not allow the innovations driving x_t (or z_t) to be serially correlated at this stage. We will discuss extensions to allow for this in section 4 where a corresponding condition will be introduced. Deo (2000) provides examples of commonly used stochastic volatility and generalised autoregressive-conditional heteroskedasticity (GARCH) processes that satisfy Assumption 1 (b).

Remark 3. For transparency, the structure in (1), (2) and (3) is exposited in terms of a scalar latent variable, z_t. However, this is without loss of generality as one may consider that $z_t = \gamma' z_t^*$ where z_t^* is a p-vector of latent variables, such that z_t satisfies the conditions stated above. □

Under Assumption 1, the conditions of Lemma 1 of Boswijk *et al.* (2015) are satisfied such that the following weak convergence result holds:

$$\left(T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\epsilon_t, T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\epsilon_s\epsilon_t'\right) \xrightarrow{w} \left(M_\eta(\cdot), \int_0^1 M_\eta(s)dM_\eta(s)'\right) \tag{4}$$

where $M_\eta(\cdot) := [M_{\eta x}(\cdot), M_{\eta z}(\cdot), M_{\eta y}(\cdot)]'$ is a Gaussian martingale satisfying

$$\begin{aligned}\begin{bmatrix} M_{\eta x}(\cdot) \\ M_{\eta z}(\cdot) \\ M_{\eta y}(\cdot)\end{bmatrix} &:= H\begin{bmatrix}\int_0^{\cdot} d_1(s)dB_1(s) \\ \int_0^{\cdot} d_2(s)dB_2(s) \\ \int_0^{\cdot} d_3(s)dB_3(s)\end{bmatrix} \\ &= \begin{bmatrix} h_{11}\{\int_0^1 d_1(s)^2\}^{1/2} & 0 & 0 \\ 0 & h_{22}\{\int_0^1 d_2(s)^2\}^{1/2} & 0 \\ h_{31}\{\int_0^1 d_1(s)^2\}^{1/2} & h_{32}\{\int_0^1 d_2(s)^2\}^{1/2} & h_{33}\{\int_0^1 d_3(s)^2\}^{1/2}\end{bmatrix}\begin{bmatrix} B_{\eta 1}(\cdot) \\ B_{\eta 2}(\cdot) \\ B_{\eta 3}(\cdot)\end{bmatrix},\end{aligned}$$

with $[B_1(\cdot), B_2(\cdot), B_3(\cdot)]'$ a 3×1 vector of independent standard Brownian motion processes and

$$B_{\eta i}(\cdot) := \{\textstyle\int_0^1 d_i(s)^2\}^{-1/2}\int_0^{\cdot} d_i(s)dB_i(s), \;\; i = 1, 2, 3.$$

We can also write $B_{\eta i}(\cdot) \stackrel{d}{=} \int_0^{\cdot} dB_i(\eta_i(s))$, $i = 1, 2, 3$, where $\eta_i(\cdot)$ denotes the variance profile

$$\eta_i(\cdot) := \{\textstyle\int_0^1 d_i(s)^2\}^{-1}\displaystyle\int_0^{\cdot} d_i(s)^2 ds, \;\; i = 1, 2, 3$$

such that $B_{\eta i}(\cdot)$ is a variance-transformed Brownian motion, i.e. a Brownian motion under a modification of the time domain; see, for example, Davidson (1994). Notice that under unconditional homoskedasticity, $\eta_i(s) = s$.

In what follows it will also prove convenient to define the two Ornstein-Uhlenbeck-type processes $B_{\eta 1, c_x}(\cdot)$ and $B_{\eta 2, c_z}(\cdot)$:

$$\begin{aligned} B_{\eta 1, c_x}(r) &:= \textstyle\int_0^r e^{-(r-s)c_x} dB_{\eta 1}(s) \\ B_{\eta 2, c_z}(r) &:= \textstyle\int_0^r e^{-(r-s)c_z} dB_{\eta 2}(s)\end{aligned}$$

for $r \in [0,1]$, along with

$$M_{\eta x, c_x}(\cdot) := h_{11}\{\textstyle\int_0^1 d_1(s)^2 ds\}^{1/2} B_{\eta 1, c_x}(\cdot)$$
$$M_{\eta z, c_z}(\cdot) := h_{22}\{\textstyle\int_0^1 d_2(s)^2 ds\}^{1/2} B_{\eta 2, c_z}(\cdot).$$

In addition, it is useful to define the innovation variance-covariance matrix in the unconditionally homoskedastic case, setting $D_t = I$ without loss of generality, as follows:

$$HH' = \begin{bmatrix} h_{11}^2 & 0 & h_{11}h_{31} \\ 0 & h_{22}^2 & h_{22}h_{32} \\ h_{11}h_{31} & h_{22}h_{32} & h_{31}^2 + h_{32}^2 + h_{33}^2 \end{bmatrix} =: \begin{bmatrix} \sigma_x^2 & 0 & \sigma_{xy} \\ 0 & \sigma_z^2 & \sigma_{zy} \\ \sigma_{xy} & \sigma_{zy} & \sigma_y^2 \end{bmatrix} =: \Omega.$$

In the context of (1), a number of possibilities exist for the predictability of y_t by the observed x_{t-1} and the unobserved z_{t-1}. One potential case that has received much attention in the literature is the predictive regression where y_t is predictable only by the observed variable x_{t-1}, so that $\beta_x \neq 0$ while $\beta_z = 0$. This forms the alternative hypothesis in the predictive regression tests discussed in section 2, where the corresponding null is that $\beta_x = 0$, and, in the context of our model, that $\beta_z = 0$ also so that y_t is unpredictable under the predictive regression null. However, it is also a possibility that y_t is predictable only by the unobserved variable z_{t-1}, with x_{t-1} playing no role in the predictability of y_t. In this case, $\beta_x = 0$ and $\beta_z \neq 0$, and any indication of predictability by x_{t-1} would be spurious. A final possibility is that $\beta_x \neq 0$ and $\beta_z \neq 0$ so that y_t is predictable by both x_{t-1} and z_{t-1}, although in this context it is not possible to estimate a correctly specified predictive regression since z_t is unobserved. We summarize these four cases using the following taxonomy of hypotheses:

$$\begin{array}{lll} H_u: & \beta_x = 0,\ \beta_z = 0 & y_t \text{ is unpredictable} \\ H_x: & \beta_x \neq 0,\ \beta_z = 0 & y_t \text{ is predictable by } x_{t-1} \\ H_z: & \beta_x = 0,\ \beta_z \neq 0 & y_t \text{ is predictable by } z_{t-1} \\ H_{xz}: & \beta_x \neq 0,\ \beta_z \neq 0 & y_t \text{ is predictable by } x_{t-1} \text{ and } z_{t-1} \end{array}$$

In hypothesis testing terms, standard predictive regression tests therefore attempt to distinguish between the null H_u and the alternative H_x. In this paper we first consider the impact of the presence of z_{t-1} in the DGP on such standard predictive regression tests, that is we investigate the behaviour of predictive regression 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 predictive regression invalidity, where the appropriate composite null is H_u or H_x, and the alternative H_z or H_{zx}.

3 Asymptotic Behaviour of Predictive Regression Tests

To fix ideas, as in Cavanagh *et al.* (1995), we first consider the basic predictive regression test of H_u against H_x, based on the t-ratio for testing $\beta_x = 0$ in the fitted linear regression

$$y_t = \hat{\alpha}_y + \hat{\beta}_x x_{t-1} + \hat{\epsilon}_{yt}, \qquad t = 1, ..., T. \tag{5}$$

This test statistic is given by

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

where

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

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

The limit distribution of t_u under Assumption 1 is shown in the next theorem.

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

$$t_u \xrightarrow{w} \frac{g_x \int_0^1 \bar{M}_{\eta x, c_x}(r)^2 + g_z \int_0^1 \bar{M}_{\eta x, c_x}(r) M_{\eta z, c_z}(r) + \int_0^1 \bar{M}_{\eta x, c_x}(r) dM_{\eta y}(r)}{\sqrt{\left\{ h_{31}^2 \int_0^1 d_1(r)^2 + h_{32}^2 \int_0^1 d_2(r)^2 + h_{33}^2 \int_0^1 d_3(r)^2 \right\} \int_0^1 \bar{M}_{\eta x, c_x}(r)^2}} \tag{6}$$

where $\bar{M}_{\eta x, c_x}(r)$ and $\bar{M}_{\eta z, c_z}(r)$ are the de-meaned versions of $M_{\eta x, c_x}(r)$ and $M_{\eta z, c_z}(r)$, respectively; that is,

$$\begin{aligned} \bar{M}_{\eta x, c_x}(r) &:= M_{\eta x, c_x}(r) - \int_0^1 M_{\eta x, c_x}(s) ds \\ \bar{M}_{\eta z, c_z}(r) &:= M_{\eta z, c_z}(r) - \int_0^1 M_{\eta z, c_z}(s) ds. \end{aligned}$$

Remark 4. Disregarding the effects of heteroskedasticity, while it is well known from Cavanagh *et al.* (1995) that the limit distribution of t_u under H_u depends on the (unknown) value of c_x whenever $\sigma_{xy} \neq 0$, the limit expression (6) also shows the dependence of t_u on g_z under H_z (where $g_x = 0$ but $g_z \neq 0$). Consequently, the use of asymptotic critical values appropriate for t_u (or the feasible versions of the test developed in Cavanagh *et al.*, 1995, and Campbell and Yogo, 2006) under H_u will not result in a size-controlled procedure 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 β_z). Under H_{xz}, where both $g_x \neq 0$ and $g_z \neq 0$, any rejection by t_u could not uniquely be ascribed to the role of x_{t-1}, potentially suggesting the existence of a well-specified predictive regression that is in fact under-specified due to the omission of the unobserved z_t. □

In addition to consideration of t_u, we also analyze the point optimal variant of this test introduced by Campbell and Yogo (2006). For a known value of ρ_x, this (infeasible) 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 $\hat{\rho}_x$ is the autoregressive coefficient estimator

$$\hat{\rho}_x := \frac{\sum_{t=1}^{T}(x_{t-1}-\bar{x}_{-1})x_t}{\sum_{t=1}^{T}(x_{t-1}-\bar{x}_{-1})^2}.$$

In the case where $s_{xy}=0$, Q and t_u coincide.

The limit distribution of Q under Assumption 1 is given below.

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

$$Q \xrightarrow{w} \frac{g_x\int_0^1 \bar{M}_{\eta x,c_x}(r)^2 + g_z\int_0^1 \bar{M}_{\eta x,c_x}(r)M_{\eta z,c_z}(r) + \int_0^1 \bar{M}_{\eta x,c_x}(r)dM_{\eta y}(r) - \frac{h_{31}}{h_{11}}\int_0^1 \bar{M}_{\eta x,c_x}(r)dM_{\eta x}(r)}{\sqrt{\left\{h_{32}^2\int_0^1 d_2(r)^2 + h_{33}^2\int_0^1 d_3(r)^2\right\}\int_0^1 \bar{M}_{\eta x,c_x}(r)^2}}. \tag{7}$$

We observe that the limit expressions in (6) and (7) are identical when $h_{31}=0$ (which is equivalent to $\sigma_{xy}=0$ in the unconditionally homoskedastic case). Given the close relationship between the t_u and Q statistics (and their corresponding limit distributions), we would again anticipate potential asymptotic size distortions under H_z. We will now proceed to investigate the extent of these distortions. Before doing so, it should be noted that other predictive regression 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 Breitung and Demetrescu (2015) and Cai *et al.* (2015). The issues we discuss in this paper are pertinent irrespective of which particular predictive regression test one uses, in cases where the putative predictor is persistent. They are also relevant for the case where a putative predictive regression contains multiple regressors (multiple predictors), see for example the Wald-based tests discussed in section 3.3 of Breitung and Demetrescu (2015), or where the putative predictive regression (and the putative predictors and the latent variables) contains a general deterministic component of the form considered in section 3.2 of Breitung and Demetrescu (2015). In the latter case, all of the distributional results given in this paper (which are for the case where a constant is included in the regression) continue to hold provided the de-meaned and tied-down Brownian-based processes which appear are appropriately re-defined to the deterministic component being considered.

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 $d_i(\cdot)=1$, $i=1,2,3$. We then simulate the limit distributions using 10,000 Monte Carlo replications, approximating the Brownian motion processes in the limiting functionals for (6) and (7) 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 is easily shown) $h_{31}^2/(h_{31}^2+h_{32}^2+h_{33}^2) = \sigma_{xy}^2/\sigma_x^2\sigma_y^2$, 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 asymptotic sizes of two-sided tests under H_z as functions of the parameter $g_z = \{0, 2.5, 5.0, ...50.0\}$ ($g_z = 0$ corresponds to size) with $g_x = 0$. For each of the four pairings $c_x = c_z = \{0, 5, 10, 20\}$ we set $\sigma_x^2 = \sigma_z^2 = \sigma_y^2 = 1$, and consider $\sigma_{xy} = \sigma_{zy} = 0$ plus $\sigma_{xy} = -0.70$ with $\sigma_{zy} = \{0, -0.35, 0.35, -0.70, 0.70\}$.[2] Setting $c_x = c_z$ is not a requirement here, but simply facilitates keeping the observed and latent predictors balanced in terms of their persistence properties. The results of this asymptotic size simulation exercise are shown in Figures 1-4.

Results for $c_x = c_z = 0$ are shown in Figure 1. We observe that the sizes of t_u and Q are growing monotonically, and quite rapidly, from the baseline level of 0.10 with increasing $g_z > 0$, thereby giving rise to an ever-increasing likelihood of ascribing spurious predictive ability to x_{t-1}. Both tests' sizes are seen throughout to exceed 0.85 for $g_z = 50$, while even a value of g_z as small as $g_z = 12.5$ always produce 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). Across the sub-figures, the size distortions appear little influenced by the value taken by σ_{zy}.

In Figure 2 the same set of simulations are conducted but now with $c_x = c_z = 5$. Qualitatively, the same comments apply here as for the case $c_x = c_z = 0$. That said, we do observe that the over-sizing now manifests itself slightly more slowly with increasing g_z. Indeed, when $\sigma_{xy} = -0.70$ and $\sigma_{zy} < 0$, some modest under-size is observed for small values of g_z. However, both sizes are still about 0.70 or higher once $g_z = 50$ so spurious predictability remains a serious issue. Figures 3 and 4 repeat the analyses with $c_x = c_z = 10$ and $c_x = c_z = 20$, respectively, and we see the extent of the spurious predictability problem continues to diminish, although it is always still very much in evidence when we consider the larger values of g_z. When $\sigma_{xy} = -0.70$ and $\sigma_{zy} < 0$, the non-monotonicity of the size profile is now more pronounced, and it is clear that, other things equal, the size distortions increase in the value of σ_{zy}. Our findings that the significance of the spurious predictability problem is inversely related to the magnitude of $c_x = c_z$ would seem to follow from intuition. As we increase the value of $c_x = c_z$, both x_{t-1} and z_{t-1} become less persistent processes and (in a very loose sense) start to have features that are more in common with stationary, rather than integrated, processes. At this point the model mis-specification begins to have a diminishing effect, being more akin to a classical mis-specification problem between stationary variables, and less like a Granger-Newbold spurious regression between (pure or local to) $I(1)$ variates problem.

Certainly then, at least for high-persistence processes, it would be difficult to argue that

[2] It is a requirement that $\sigma_{xy}^2 + \sigma_{zy}^2 < 1$ (given $\sigma_x^2 = \sigma_z^2 = \sigma_y^2 = 1$) in order to ensure Ω is positive definite.

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. Although we have focussed this analysis on OLS-based predictive regression tests, similar qualitative results will pertain for other predictive regression tests including the recently proposed IV-based tests of Breitung and Demetrescu (2015) 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 latent 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 predictive regression tests, we should always remain alert to the possibility of spurious predictability.

4 A Stationarity Test for Predictive Regression Invalidity

Given the potential for standard predictive regression 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 predictive regression tests based on x_{t-1} are valid. Rejection, however, would indicate the presence of an unobserved z_{t-1} component in the generating process for y_t, thereby signalling the invalidity of predictive regression tests based on x_{t-1}. Formally, then, we wish to test the null hypothesis that $\beta_z = 0$, i.e. H_u or H_x, against the alternative that $\beta_z \neq 0$, i.e. H_z or 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 well-known KPSS test; see also Nyblom (1989). Consider first the KPSS-type statistic for serially independent errors applied to the residuals $\hat{\epsilon}_{yt}$ from (5):

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

where s_y^2 is as defined previously. When $\beta_z \neq 0$, the residuals $\hat{\epsilon}_{yt}$ from (5) incorporate the omitted $\beta_z z_{t-1}$ term in (1), hence the persistence in z_{t-1} is passed to $\hat{\epsilon}_{yt}$, and a test of $\beta_z = 0$ against $\beta_z \neq 0$ can be formed as a test for stationarity of $\hat{\epsilon}_{yt}$, rejecting for large values of S. Specifically, assume that $c_z = 0$ and consider rewriting (1) as

$$y_t = \alpha_y + \beta_x x_{t-1} + r_{t-1} + \epsilon_{yt}, \qquad r_t = r_{t-1} + u_t, \qquad u_t = \beta_z \epsilon_{zt}. \tag{8}$$

Then it is clear that testing $\beta_z = 0$ against $\beta_z = g_z T^{-1}$ in (1) is precisely the same problem as testing $V(u_t) =: \sigma_u^2 = 0$ against $\sigma_u^2 = (g_z T^{-1})^2$ in (8), with $g_z = 0$ under both null hypotheses (that is, under $\beta_z = 0$ and under $\sigma_u^2 = 0$). Now, if we temporarily assume that x_t is strictly exogenous and ϵ_{yt} and ϵ_{zt} to be independent IID normal random variates, then the test which rejects for large values of S can be shown to be the locally best invariant (to α_y, α_x, α_z, β_x and σ_y^2) test of the null $\sigma_u^2 = 0$ against the local alternative $\sigma_u^2 = g_z^2 T^{-2}$ in (8). Hence a KPSS-type test is relevant for our testing problem where we seek to distinguish between $\beta_z = 0$ and $\beta_z \neq 0$. Of course, in our more general model we do not impose $c_z = 0$ (nor the other temporary assumptions listed above), so in these more general circumstances we may reasonably 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 (5). As such, a rejection by this test indicates that the fitted regression in (5) is not a valid predictive regression. As with the failure of any mis-specification test, this does not tell us why the regression has failed. We do know that S is designed to deliver a test which is (approximately) locally optimal in the direction of z_{t-1} being an omitted predictor, but a rejection does not mean that x_{t-1} is not a valid predictor for y_t. Indeed, z_{t-1} might reasonably be viewed as a proxy for more general mis-specification in the underlying regression model. An obvious example is provided by the case where the true slope coefficient in (5) displays time-varying behaviour, such as has been considered in, for example, Paye and Timmermann (2006) and Cai *et al.* (2015). It is therefore important to stress that our proposed test is one for the invalidity of the putative predictive regression, not as a test for the invalidity of the putative predictor, x_{t-1}.

Because it is crucial in the present setting to account for the possibility of correlation between ϵ_{xt} and ϵ_{yt} ($h_{31} \neq 0$), we do not use the simple form of the S statistic given above, but follow Shin (1994) by including an additional regressor Δx_t in the regression used to construct the KPSS-type statistic. That is, in place of (5) we use the fitted linear regression

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

and construct S using the residuals $\hat{e}_t$ from (9), thereby redefining S as

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

where $s^2 := (T-3)^{-1} \sum_{t=1}^{T} \hat{e}_t^2$.

In Theorem 3 we now detail the limiting distribution of S under Assumption 1. This is followed by some remarks concerning the result in the theorem and an extension to allow for serial dependence in the innovations driving x_{t-1}.

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

$$S \xrightarrow{w} \{h_{32}^2 \textstyle\int_0^1 d_2(r)^2 + h_{33}^2 \int_0^1 d_3(r)^2\}^{-1} \int_0^1 \{F(r, c_x) + g_z G(r, c_x, c_z)\}^2 \tag{10}$$

where

$$\begin{aligned} F(r, c_x) &:= \mathbb{B}_\eta^*(r) - \textstyle\int_0^r \bar{B}_{\eta 1, c_x}(s) \{\int_0^1 \bar{B}_{\eta 1, c_x}(s)^2\}^{-1} \int_0^1 \bar{B}_{\eta 1, c_x}(s) dB_\eta^*(s) \\ G(r, c_x, c_z) &:= h_{22} \{\textstyle\int_0^1 d_2(s)^2\}^{1/2} \left[\int_0^r \bar{B}_{\eta 2, c_z}(s) - \int_0^r \bar{B}_{\eta 1, c_x}(s) \{\int_0^1 \bar{B}_{\eta 1, c_x}(s)^2\}^{-1} \int_0^1 \bar{B}_{\eta 1, c_x}(s) B_{\eta 2, c_z}(s) \right] \end{aligned}$$

with: $\mathbb{B}^*_\eta(r) := B^*_\eta(r) - rB^*_\eta(1)$ *the tied-down version of* $B^*_\eta(r)$*;* $\bar{B}_{\eta 1,c_x}(r) := B_{\eta 1,c_x}(r) - \int_0^1 B_{\eta 1,c_x}(s)$ *and* $\bar{B}_{\eta 2,c_z}(r) := B_{\eta 2,c_z}(r) - \int_0^1 B_{\eta 2,c_z}(s)$ *the de-meaned versions of* $B_{\eta 1,c_x}(r)$ *and* $B_{\eta 2,c_z}(r)$*, respectively; and, finally,* $B^*_\eta(r) := h_{32}\{\int_0^1 d_2(s)^2\}^{1/2}B_{\eta 2}(r) + h_{33}\{\int_0^1 d_3(s)^2\}^{1/2}B_{\eta 3}(r)$*.*

Remark 5. The limit expression in (10) clearly shows how g_z enters the asymptotic distribution of S under H_z and H_{xz}. In particular, it is the presence of the term $g_z G(r, c_x, c_z)$ that is seen to be the source of power for the test based on S to distinguish between H_u or H_x and H_z or H_{xz} (noting that, by construction, S is exact invariant to β_x). Notice that the distribution in (10) does not depend on h_{31}, and, hence, does not depend on the correlation between ϵ_{xt} and ϵ_{yt}.

Remark 6. Under H_u or H_x, where $g_z = 0$, the limit distribution in (10) simplifies to $\int_0^1 \left[\{h_{32}^2\int_0^1 d_2(r)^2 + h_{33}^2\int_0^1 d_3(r)^2\}^{-1/2}F(r,c_x)\right]^2$. Notice, in this case we may write

$$\begin{aligned}\{h_{32}^2\textstyle\int_0^1 d_2(r)^2 + h_{33}^2\int_0^1 d_3(r)^2\}^{-1/2}F(r,c_x) &= D_\eta(r) - rD_\eta(1) \\ &\quad - \textstyle\int_0^r \bar{B}_{\eta 1,c_x}(s)\{\int_0^1 \bar{B}_{\eta 1,c_x}(s)^2\}^{-1}\int_0^1 \bar{B}_{\eta 1,c_x}(s)D_\eta(s)\end{aligned}$$

where $D_\eta(r) := \{h_{32}^2\int_0^1 d_2(s)^2 + h_{33}^2\int_0^1 d_3(s)^2\}^{-1/2}[h_{32}\{\int_0^1 d_2(s)^2\}^{1/2}B_{\eta 2}(r) + h_{33}\{\int_0^1 d_3(s)^2\}^{1/2}B_{\eta 3}(r)]$ is a standardised heteroskedastic Brownian motion that is independent of $B_{\eta 1}(r)$. Consequently, where $g_z = 0$, the limit distribution of S depends only on c_x and any unconditional heteroskedasticity present in ϵ_t.

Remark 7. In our analysis we have assumed that the increments of the process for x_t, i.e. ϵ_{xt}, are serially uncorrelated, by virtue of e_t being a martingale difference sequence. More generally we might consider a linear process assumption for ϵ_{xt} of the form

$$\epsilon_{xt} = \sum_{i=0}^{\infty} \theta_i v_{x,t-i} \tag{11}$$

where $v_{x,t}$ denotes 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 ϵ_{yt} remains uncorrelated with the increments of x_t at all lags (i.e. x_t is weakly exogenous with respect to ϵ_{yt}) under this structure. In this case, it may be shown that the limiting results given in Theorem 3 above and in Theorems 4-6 which follow will continue to hold provided we replace (9) in the calculation of S with the augmented variant

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

where p satisfies the standard rate condition that $1/p + p^3/T \to 0$, as $T \to \infty$, and where it is assumed that $T^{1/2}\sum_{i=p+1}^\infty |\delta_i| \to 0$, where $\{\delta_i\}_{i=1}^\infty$ are the coefficients of the $AR(\infty)$ process obtained by inverting the $MA(\infty)$ process in (11). Similarly to Breitung and Demetrescu (2015), we would also need to restrict the amount of serial dependence allowed in the conditional variances via the cross-product moment assumption that $\sup_{i,j\geq 1}\|\tau_{ij}\| < \infty$, where $\tau_{ij} :=$

$E(e_t e_t' \otimes e_{t-i} e_{t-j}')$, with $\otimes$ denoting the Kronecker product. Serial correlation of a similar form in the increments of the unobserved process z_t, i.e. ϵ_{zt}, will have no impact on our large sample results under the null hypothesis, H_u/H_x, although an effect would be apparent under H_z/H_{xz}. As is standard in the predictive regression literature, we maintain the assumption that ϵ_{yt} is serially uncorrelated, which is why, unlike in the setting considered in Shin (1994), we need only include lags of Δx_t, rather than both leads and lags thereof.

Remark 8. Consider again the discussion just before the start of section 3.1 relating to the case where the putative predictive regression contains multiple regressors and/or more general deterministic components. These extensions can easily be handled in the context of our proposed predictive regression 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 Breitung and Demetrescu (2015), 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 (and its bootstrap analogue, S^*, given below) using the residuals $\hat{e}_t$ from the regression of y_t (y_t^* for the bootstrap S^* statistic) onto $\mathbf{f}_t$, $\mathbf{x}_{t-1}$ and $\Delta\mathbf{x}_{t-1}$ (and lags of $\Delta\mathbf{x}_{t-1}$ in the case considered in Remark 7). Doing so would alter the form of the limit distributions given in Theorem 3 and in the sequel, as noted earlier, 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 3 is therefore that if we wish to base a test for predictive regression invalidity on S, then we need to address the fact that when we treat H_u or H_x as the null hypothesis, the limit null distribution of S is not pivotal. In order to account for the dependence of the limit distribution of S 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 $x := [x_0, x_1, ..., x_T]'$ as a fixed regressor when implementing the bootstrap procedure. The next subsection details this procedure.

4.2 A Fixed Regressor Wild Bootstrap Stationarity Test

A conventional approach to obtaining wild bootstrap critical values with which to compare 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 turns out to be 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 a standard bootstrap sample 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 the bootstrap sample data for x_t from bootstrap samples of e_{xt} (a wild bootstrap re-sampling scheme from Δx_t for example) would require knowledge of c_x. Since we assume c_x is not known, recourse would then be made to its estimation. However, it is well-known that conventional estimators of c_x (such as the one based on a regression of x_t on x_{t-1}) are not consistent and if such estimators were employed in constructing bootstrap samples of x_t, the large sample distribution of the resulting bootstrap analogue of S would be different to that of S under H_u/H_x, with the consequence that the resulting bootstrap test would not be correctly sized, even asymptotically. To avoid this problem, we abstract away from estimation of c_x altogether and instead follow the approach taken in Hansen (2000), considering a bootstrap procedure which uses $x := [x_0, x_1, ..., x_T]'$ 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 the original KPSS-type statistic, S.

We now outline the steps involved in our proposed fixed regressor wild bootstrap, collected together in Algorithm 1.

Algorithm 1 (Fixed Regressor Wild Bootstrap):

(i) Construct the wild bootstrap innovations $y_t^* := \hat{e}_t w_t$, where w_t, $t = 1, \ldots, T$, is an $IID\, N(0,1)$ sequence.

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

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

where $\hat{\epsilon}_{yt}^*$ are the OLS residuals from the fitted regression

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

and where $(s_y^*)^2 := (T-2)^{-1} \sum_{t=1}^{T} (\hat{\epsilon}_{yt}^*)^2$.

(iii) Define the corresponding p-value as $P_T^* := 1 - G_T^*(S)$ with $G_T^*(\cdot)$ denoting the conditional (on the original data) cumulative density function (cdf) of S^*.

(iv) The wild bootstrap test of H_u/H_x at level ξ rejects in favour of H_z/H_{xz} if $P_T^* \leq \xi$.

Remark 9. In practice, the cdf $G_T^*(\cdot)$ required in step (iii) of Algorithm 1 will be unknown, but can be approximated in the usual way through numerical simulation. This is achieved by generating B (conditionally) independent bootstrap statistics, S_k^*, $k = 1, ..., B$, each computed as in Algorithm 1 above, but from $y_{t,k}^* := \hat{e}_t w_{t,k}$, $t = 1, ..., T$, with $\{\{w_t\}_{t=1}^T\}_{k=1}^B$ a doubly independent $N(0,1)$ sequence. The simulated bootstrap p-value for S is then computed as $\tilde{P}_T^* := B^{-1} \sum_{k=1}^{B} \mathbb{I}(S_k^* > S)$, and is such that $\tilde{P}_T^* \to P_T^*$ almost surely as $B \to \infty$. The choice of B is discussed by, *inter alia*, Davidson and MacKinnon (2000). Note that an asymptotic standard error for $\tilde{P}_T^*$ is given by $[\tilde{P}_T^*(1 - \tilde{P}_T^*)/B]^{1/2}$; cf. Hansen (1996, p.419). Finally, we may also define the associated ξ level empirical bootstrap critical value, denoted $cv_{\xi,B}$, to be the upper tail ξ percentile from the order statistic formed from the B bootstrap statistics S_k^*, $k = 1, ..., B$.

Remark 10. The use of x_{t-1} as a fixed regressor in the construction of the bootstrap KPSS-type statistic, S^*, is made explicit in (13). It is then seen that each of the bootstrap S_k^*, $k = 1, ..., B$, statistics calculated as described in Remark 9 uses the same x_{t-1} as the regressor in (13). As we shall see in section 4.3, this has important ramifications for the methods needed to prove the asymptotic validity of our proposed bootstrap procedure. In particular, this will entail the necessity to develop a conditional version of the invariance principle given in 4, jointly with a bootstrap counterpart of this result, both of which appear to be new to the literature.

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 $\hat{e}_t$, y_t^* is independent over time with zero mean and variance $\hat{e}_t^2$.

Remark 12. Although $\hat{e}_t$ depends on g_z under H_z or H_{xz}, we will show in the next subsection that this does not translate into large sample dependence of S^* on g_z. Notice also that we do not need to include Δx_t as an additional regressor, or lags thereof in the case considered in Remark 7, in (13). This is because the $\hat{e}_t$ used to construct y_t^* are free of any effects arising from the correlation between ϵ_{xt} and ϵ_{yt}, or from any weak dependence in ϵ_{xt}.

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, the distribution of S^* given the data converges weakly to the random distribution which obtains by conditioning the limit in (10) on the weak limit B_1 of $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot \rfloor} e_{1t}$. We also show that under H_u/H_x the distribution of the test statistic S conditional on x converges weakly to the same random distribution, which in turn allows us to establish the asymptotic validity of our bootstrap test. The asymptotic results we provide will therefore be likely to have wider applicability in other scenarios where a fixed regressor bootstrap is used with (near-) integrated regressors.

It is a well known result that even if a random sequence, say (X_T, Y_T), converges to some (X, Y) in a strong sense (e.g., almost surely), the conditional distribution of X_T given Y_T need not converge to the conditional distribution of X given Y. Consequently, Theorem 3, where the limit distribution of S is established, cannot be taken to imply that S conditional on x converges weakly to the limit in (10) conditioned on B_1. Nevertheless, it is not unreasonable to expect that this result holds true, and here we develop the necessary theory in order to formally prove it is in fact so.

Analogously to Theorem 3, whose validity is based on an invariance principle, a conditional and a bootstrap version of that theorem can be based on a *conditional* joint invariance principle for the original and the bootstrap data. We are not aware of the existence of such a result in the literature, and so our first step here is to establish it. In order to achieve this we 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 σ-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 σ-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 13. 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 Ω_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 $E(\varepsilon_{it}|\varepsilon_{1t}, \varepsilon_{2t}^2, \varepsilon_{3t}^2) = 0$, $i = 2, 3$. 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 leading 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 ε_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 the section of Gonçalves and Kilian (2004) that deals with wild bootstrap testing. 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 4 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 4 we now establish two results: 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, Denken and Gordin, 2003, Crimaldi and Pratelli, 2005), and, second, a bootstrap extension of that result. Besides their importance for obtaining our main result, the conclusions of Theorem 4 serve as a vehicle to discuss the meaning of bootstrap validity in the context of weak convergence to random measures. Analogously to the definition of the vector x, let $y := [y_1, y_2, ..., y_T]'$ and $z := [z_0, z_1, ..., z_T]'$.

Theorem 4. *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 $B^{\dagger}(\cdot) := [B_1^{\dagger}(\cdot), B_2^{\dagger}(\cdot), B_3^{\dagger}(\cdot)]'$ denote a standard trivariate Brownian motion, independent of $B(\cdot)$. Under Assumption 2, the following converge* *jointly as $T \to \infty$:*

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

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

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

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

Remark 14. Let $E_x(\cdot) := E(\cdot|x)$ and $E^*(\cdot) := E(\cdot|x, y, z)$, the latter denoting expectation given the data in standard bootstrap notation. The definition of the joint weak convergence of random measures result established in Theorem 4, is that for all bounded continuous real functions f and g on $\mathcal{D}^3$ and $\mathcal{D}^4$, respectively, it holds that

$$\begin{bmatrix} E_x(f(U'_{\lfloor T\cdot\rfloor})) \\ E^*(g(U_{\lfloor T\cdot\rfloor 1}, U'_{\lfloor T\cdot\rfloor b})) \end{bmatrix} \xrightarrow{w} \begin{bmatrix} E\left(f\left(B'\right)\middle| B_1\right) \\ E(g(B_1, (B^\dagger)')\big| B_1) \end{bmatrix}$$

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

Remark 15. An implication of Theorem 4 is that if a continuous bounded function f is used to define a statistic $\phi = f(U'_{\lfloor T\cdot\rfloor})$, and the 'fixed-regressor bootstrap' statistic $\phi^* = f(U_{\lfloor T\cdot\rfloor 1}, U_{\lfloor T\cdot\rfloor b,2}, U_{\lfloor T\cdot\rfloor b,3})$ is used to construct a distributional approximation for ϕ, then ϕ conditional on x and ϕ^* conditional on the data will jointly converge weakly to the same random measure, defined by $f\left(B'\right)|B_1$. As a consequence, the bootstrap approximation is consistent in the following sense: it holds that

$$\sup_{u\in\mathbb{R}} |P_x\left(\phi \le u\right) - P^*(\phi^* \le u)| \xrightarrow{p} 0 \tag{14}$$

provided that the cumulative distribution process of $f\left(B'\right)\big| B_1$ is a.s. 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 ϕ^* conditional on the data consistently estimates the large-sample distribution of the original statistic ϕ conditional on the 'fixed regressor' x. This result is more general than the usual formulation of bootstrap validity, where either an unconditional $P\left(\phi \le u\right)$ appears or a conditional $P_x\left(\phi \le u\right)$ with a non-random limit. □

In order to obtain the analogue of (14) for S and S^*, we need to discuss limits involving some stochastic integrals. Because these are not continuous transformations, the discussion in Remark 15 does not apply directly and so we also need the result given next in Theorem 5. Here the dimension of the bootstrap partial-sum process is reduced to one and the process itself is constructed from delta-measurable quantities $\tilde{e}_{Tt}$ that we shall subsequently specify to be the residuals $\hat{e}_t$ from the regression in (9).

Theorem 5. *Let $\tilde{e}_{Tt}$ $(t = 1, ..., T)$ be scalar measurable functions of the data x, y, z and such that $\sum_{t=1}^{\lfloor Tr\rfloor} \tilde{e}^2_{Tt} \xrightarrow{p} \int_0^r m^2(s)ds$ for all $r \in [0, 1]$, where $m(\cdot)$ is a square-integrable real function on*

$[0,1]$. *Introduce* $\tilde{\epsilon}_{tb} := w_t \tilde{e}_{Tt}$, $\tilde{U}_{tb} := T^{-1/2} \sum_{s=1}^{t} \tilde{\epsilon}_{sb}$ *(*$t = 1, ..., T$*), and* $\tilde{B}_m^{\dagger}(\cdot) := \int_0^{\cdot} m(s) dB_1^{\dagger}(s)$*, where* $B_1^{\dagger}$ *is as in Theorem 4. Under Assumption 2, the following converge jointly as* $T \to \infty$*:*

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

in the sense of weak convergence of random measures on $\mathcal{D}^3 \times \mathbb{R}^2$*, and*

$$\left(U_{\lfloor T\cdot \rfloor 1}, \tilde{U}_{\lfloor T\cdot \rfloor b}, T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{t-1} \epsilon_{xs} \tilde{\epsilon}_{tb} \right) \Bigg| x, y, z \xrightarrow{w} \left(B_1(\cdot), \tilde{B}_m^{\dagger}(\cdot), \int_0^1 M_{\eta x}(s) d\tilde{B}_m^{\dagger}(s) \right) \Bigg| B_1$$

in the sense of weak convergence of random measures on $\mathcal{D}^2 \times \mathbb{R}$*.*

We are now in a position to establish in Theorem 6 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 6. *Under DGP (1)-(3) and Assumption 2, the following converge jointly as* $T \to \infty$*, in the sense of weak convergence of random measures on* $\mathbb{R}$*:*

$$S|x \xrightarrow{w} \{h_{32}^2 \textstyle\int_0^1 d_2(r)^2 + h_{33}^2 \int_0^1 d_3(r)^2\}^{-1} \int_0^1 \{F(r, c_x) + g_z G(r, c_x, c_z)\}^2 \Big| B_1 \tag{15}$$

and

$$S^*|x, y, z \xrightarrow{w} \{h_{32}^2 \textstyle\int_0^1 d_2(r)^2 + h_{33}^2 \int_0^1 d_3(r)^2\}^{-1} \int_0^1 F^{\dagger}(r, c_x)^2 \Big| B_1, \tag{16}$$

where

$$F^{\dagger}(r, c_x) := B_\eta^{\dagger *}(r) - r B_\eta^{\dagger *}(1) - \textstyle\int_0^r \bar{B}_{\eta 1, c_x}(s) \{\int_0^1 \bar{B}_{\eta 1, c_x}(s)^2\}^{-1} \int_0^1 \bar{B}_{\eta 1, c_x}(s) dB_\eta^{\dagger *}(s)$$

with

$$B_\eta^{\dagger *}(r) := h_{32} \{\textstyle\int_0^1 d_2(s)^2\}^{1/2} B_{\eta 2}^{\dagger}(r) + h_{33} \{\int_0^1 d_3(s)^2\}^{1/2} B_{\eta 3}^{\dagger}(r).$$

Remark 16. A comparison of (15) and (16) in Theorem 6 shows that the bootstrap statistic S^*, conditional on the data, and the original statistic S, conditional on x, share the same first-order asymptotic distribution when $g_z = 0$; that is, under the null hypothesis, H_u/H_x. An implication of this result, formalised in Corollary 1 below, is that it enables us to establish the asymptotic validity of the bootstrap test based on S^*. As usual, it is formulated in terms of bootstrap P-values. □

Corollary 1. *Under* H_u/H_x *and Assumption 2,*

$$P_T^* = P^*(S^* \leq S) \xrightarrow{w} U[0, 1] \tag{17}$$

as $T \to \infty$*.*

Remark 17. For the bootstrap statistic, S^*, the convergence result in (16) does not depend on the value of g_z; that is, the same limiting distribution is obtained under the alternative hypothesis, H_z/H_{xz}, as under the null hypothesis. In contrast, for the case of S (conditional on x), a stochastic offset, arising from the term $g_z G(r, c_x, c_z)$, is seen in the limiting distribution in (15). It is important to note, however, that the limiting distribution given in (15) does not coincide with that given in (10) in Theorem 3. The implication of this is that a bootstrap test based on S^* will have non-trivial asymptotic local power under H_z/H_{xz}, but that this will not coincide (for a given alternative) with the asymptotic local power of an (infeasible) version of S based on knowledge of the unknown parameter, c_x. □

The practical implication of the foregoing results is that comparison of the original statistic S with an empirical bootstrap critical value based on B bootstrap replications, $cv_{\xi,B}$ as defined in Algorithm 1 and Remark 9, will result in a bootstrap test that has correct asymptotic size (ξ) under H_u/H_x, and non-trivial local power under H_z/H_{xz}, the latter because while the empirical bootstrap critical value $cv_{\xi,B}$ will remain unchanged (at least in the limit), the corresponding critical value from the distribution of S conditional on x will depend on g_z. In what follows, as a matter of shorthand notation, we will denote by S_B the fixed regressor wild bootstrap procedure outlined in Algorithm 1 and Remark 9, based on B bootstrap replications, whereby S is compared to the empirical bootstrap critical value $cv_{\xi,B}$.

4.4 Asymptotic Local Power of Stationarity Tests under H_z

We now turn to a consideration of the asymptotic local power of S and S_B. We use the same set of unconditionally homoskedastic simulation models as for the size of t_u and Q in Figures 1-4, so we overlay the power information on them. For the asymptotic power of S under H_z we use the limit expression (10), having first obtained 0.10-level critical values from simulating (10) under $g_z = 0$. Since these critical values depend on knowledge of c_x, we can consider S an infeasible test against which to benchmark the power of the fixed regressor wild bootstrap procedure, 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 $cv_{\xi,B}$ (see Remark 9) with $\xi = 0.10$ in each Monte Carlo replication. For each replication, the simulated limit bootstrap critical value is obtained by simulating the limit (16) using $B = 2000$ bootstrap replications, conditioning on the simulated B_1 for that Monte Carlo replication.

From Figure 1, where $c_x = c_z = 0$, 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, we see that it follows a very similar power profile to that of S; indeed, its power marginally exceeds that of S. It is of course anticipated from the discussion in Remark 17 that S_B would not have the same asymptotic local power function as S, but the fact that its power exceeds that of S is a most welcome finding especially as S_B, unlike S, represents a feasible procedure. Similar comments apply to Figure 2, where $c_x = c_z = 5$, although a non-monotonicity in the power profiles of S and S_B is apparent for $\sigma_{xy} = -0.70$ and $\sigma_{zy} < 0$ for small g_z, with power dipping below size.

In Figures 3 and 4 (which examines the larger values of $c_x = c_z$) the powers of S and S_B appear near identical, but at a lower level than in Figures 1 and 2. The issue of non-monotonicity with power below size when $\sigma_{xy} = -0.70$ and $\sigma_{zy} < 0$ is more apparent here, revealing itself for small to moderate values of g_z. For larger values of g_z, power increases with g_z, as seen for all values of g_z when $\sigma_{xy} = -0.70$ and $\sigma_{zy} \geq 0$, and when $\sigma_{xy} = \sigma_{zy} = 0$.

The important comparison next is between the asymptotic power of S_B (restricting our attention to the feasible procedure) and the size of t_u and Q (for discussion purposes we treat these as a pair because their size profiles are qualitatively similar). From Figure 1, it is clear that when $c_x = c_z = 0$, the power of S_B exceeds the size of t_u/Q, hence the invalidity of the predictive regression is detected with greater frequency than t_u/Q spuriously reject in favour of predictability of y_t by x_{t-1}. This demonstrates the capability of S_B to detect predictive regression invalidity in cases where the important size problems associated with t_u/Q are apparent. That the power of S_B exceeds the size of t_u/Q under H_z is largely 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 predictive regression tests t_u and Q. In Figure 2, where $c_x = c_z = 5$, we again see that the power of S_B generally out-strips the sizes of t_u/Q, with the size/power differences appearing even more marked than for $c_x = c_z = 0$. The only exception to this is for $\sigma_{xy} = -0.70$ and $\sigma_{zy} < 0$ when g_z is small; however, since there is no discernible over-size for t_u/Q here, this is of little concern. Similar remarks apply to Figures 3 and 4 for the cases of larger $c_x = c_z$, with S_B power generally exceeding t_u/Q size, with the exception of the non-monotonicity region when $\sigma_{xy} = -0.70$ and $\sigma_{zy} < 0$; once again, however, we generally observe that where S_B has power below size, t_u and Q do not over-reject. Comparing across the figures, as the persistence of x_t and z_t decrease, so do the powers of S and S_B, in line with the decreasing size of t_u/Q discussed above. This is as would be expected given that, other things equal, the influence of a less persistent z_t becomes harder to detect.

5 Finite Sample Size and Power under H_z

In this section we evaluate the finite sample size properties of the predictive regression tests and the size and power of the newly proposed test for predictive regression invalidity. For the predictive regression tests, we first consider the feasible version of the t_u test proposed by Cavanagh *et al.* (1995) and the feasible version of the Q statistic proposed by Campbell and Yogo (2006), which both rely on Bonferroni bounds to control size.[3] In addition, we consider the preferred IV-based test of Breitung and Demetrescu (2015) which combines a fractional instrument with a sine function instrument, denoted IV_{comb} hereafter, comparing this against its asymptotic $\chi^2(1)$ critical value. For the predictive regression invalidity test we report results for the feasible bootstrap test S_B, based on $B = 499$ bootstrap replications.

To begin, we continue to abstract from heteroskedasticity and consider finite sample DGPs for the same settings as were used in the asymptotic simulations above. Specifically, we simulate

[3]We are grateful to Campbell and Yogo for making their Gauss code available for these two procedures.

the DGP (1)-(3) for $T = 200$ with $\alpha_y = \alpha_x = \alpha_z = 0$, $\beta_x = 0$, $s_{x,0} = s_{z,0} = 0$, $d_{it} = 1$ $(i = 1, 2, 3)$, and $e_t \sim IID\,N(0, I_3)$. Figures 5-8 report the finite sample analogues of Figures 1-4, i.e. rejection frequencies of nominal 0.10-level (two-sided for t_u, Q and IV_{comb}) tests under H_z for the same settings of g_z, $c_x = c_z, \sigma_x^2 = \sigma_z^2 = \sigma_y^2$, σ_{xy} and σ_{zy}. The simulations are again conducted using 10,000 Monte Carlo replications. On comparing Figures 5-8 with their large sample counterparts in Figures 1-4, 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 in part a consequence of the feasible tests not having the same large sample properties as the infeasible tests. The general observations made on the basis of the asymptotic simulations apply equally here, that is a feature of increasing finite sample size of the predictive regression tests as g_z increases, again suggesting an increasing likelihood of spurious predictive ability. As anticipated in the discussion of section 3.1, a similar pattern of rejections is found for IV_{comb}, where the spurious rejection frequencies are seen to be close to those associated with t_u and Q. As regards S_B, its finite sample power is found to also increase with g_z, with the invalidity of the predictive regression generally detected with greater frequency than the predictive regression tests' spurious rejections. The capability of S_B to detect predictive regression invalidity in cases where well-known predictive regression tests suffer problematic over-size is consequently also displayed in finite samples.

As a matter of practical interest, it is valuable to analyse the interplay between the occurrence of spurious rejections by the predictive regression tests and the likelihood of detecting predictive regression invalidity. To this end, we evaluate the performance of a putative two-step procedure, whereby in a first stage, the S_B test for predictive regression invalidity is applied as a pre-test, and then a given predictive regression test (t_u, Q or IV_{comb}) is only applied as a second stage if S_B fails to reject. As such, we are gauging the efficacy of using the S_B test to reduce the degree of predictive regression test over-size by pre-screening for predictive regression invalidity. In Figures 5-8 we additionally report the rejection frequencies for such two-step pre-test-based procedures, denoted by t_u^{pre}, Q^{pre} and IV_{comb}^{pre}; here, a rejection in favour of predictability of y_t by x_{t-1} is only returned if S_B non-rejects *and* the appropriate predictive regression test rejects (t_u, Q or IV_{comb}). We observe that the substantial over-size seen for the t_u, Q and IV_{comb} tests is dramatically reduced by prior application of S_B as a pre-test. Indeed, the t_u^{pre}, Q^{pre} and IV_{comb}^{pre} sizes converge to zero as g_z becomes large, driven by the power of S_B increasing in g_z. For smaller g_z we see that the t_u^{pre}, Q^{pre} and IV_{comb}^{pre} rejection frequencies rarely exceed the nominal 0.10 level, particularly for the larger values of $c_x = c_z$ considered, and in the cases where some over-sizing does remain (mostly when $c_x = c_z = 0$), the maximum size seen across g_z is around 0.18, though often much less. Comparing the original predictive regression tests with their pre-test-based counterparts, the role of the bootstrap predictive regression invalidity test S_B would appear indisputable.

We next consider the impact of unconditional heteroskedasticity in the DGP, investigating the size of IV_{comb} (and IV_{comb}^{pre}) and the size and power of S_B when the error processes are subject to a single break in volatility.[4] Specifically, we again simulate the DGP (1)-(3) for $T = 200$ with $\alpha_y = \alpha_x = \alpha_z = 0$, $\beta_x = 0$, $s_{x,0} = s_{z,0} = 0$, and $e_t \sim IID\, N(0, I_3)$, but now we let

$$d_{it} = \begin{cases} 1 & t \leq \lfloor \tau T \rfloor \\ \sigma_i & t > \lfloor \tau T \rfloor \end{cases}, \qquad i = 1, 2, 3$$

with $\tau = \{0.3, 0.7\}$ thereby allowing for two common volatility break timings, and $\sigma_i = \{1, 4, \frac{1}{4}\}$ allowing for both upward and downward volatility shifts, with the chosen magnitudes being substantial for illustrative purposes. We consider $c_x = c_z = \{0, 5, 10\}$ and for simplification abstract from time-varying correlation between ϵ_{xt}, ϵ_{zt} and ϵ_{yt} by setting $h_{31} = h_{32} = 0$. Table 1 reports results for the size of S_B, IV_{comb} and IV_{comb}^{pre} where $g_z = 0$ (note that the settings for c_z and σ_2 are irrelevant here). Tables 2 and 3 then report results for $g_z = 25$ and $g_z = 50$ respectively, where the table entries correspond to power for S_B and size for IV_{comb} and IV_{comb}^{pre}. As before, the results are for nominal 0.10-level tests, two-sided in the case of IV_{comb} and IV_{comb}^{pre}.

Turning first to Table 1, it is clear that the size of S_B is very well controlled across all the patterns of time-varying volatility of ϵ_{xt} and ϵ_{yt} we consider. 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.[5] The IV_{comb} test also displays a good degree of robustness to heteroskedasticity, although size can be a little inflated for some settings. As would be expected, IV_{comb}^{pre} displays empirical size slightly lower than that of IV_{comb}.

Finally, in Tables 2 and 3 we again see that, for a given heteroskedasticity setting, the power of S_B is increasing in g_z. However, it is clear that the presence of (unconditional) heteroskedasticity can have a substantial influence on the level of power attainable. Other things equal, a volatility increase in ϵ_{zt} (an increase in σ_2) leads to higher S_B power, with a volatility decrease in ϵ_{zt} having the opposite effect, while volatility changes in ϵ_{yt} have the reverse effect, with an increase (decrease) in σ_3 resulting in lower (higher) power for S_B. Volatility changes in ϵ_{xt} (changes in σ_1) appear to have relatively little effect. While the absolute powers can vary across the timing of the volatility changes, the directional relationships are the same for $\tau = 0.3$ and $\tau = 0.7$. 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). As in the conditionally homoskedastic case, we generally see that the power of S_B exceeds the size of IV_{comb}, although under the large volatility changes that we consider, there are a small number of cases where this ranking is reversed and S_B power falls below IV_{comb} size, typically associated with a reduction in ϵ_{zt} volatility and an increase in ϵ_{yt} volatility. In terms of the interplay between S_B and IV_{comb}, we again see that the

[4]We do not consider t_u and Q here since these procedures are not designed to be robust to heteroskedastic errors.

[5]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 nominal size observed.

pre-screened IV_{comb}^{pre} procedure always achieves a reduction in the over-size of IV_{comb}. In most cases, the reductions are substantial, as in the corresponding homoskedastic cases, resulting in an IV_{comb}^{pre} size at or below the nominal 0.10 level. In those situations where S_B has power at a similar or lower level than the IV_{comb} size, the rejection frequency of IV_{comb}^{pre} can be above the nominal level, sometimes to a greater extent than was seen under homoskedasticity, but in all cases the degree of over-size is still markedly lower than that of the original IV_{comb} test which makes no use of the predictive regression invalidity test S_B. 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 Illustration using U.S. Stock Index Returns

To illustrate how our proposed procedure may be used in practice, we consider predictability of the monthly U.S. S&P returns, i.e. the differenced log S&P stock price index, using either the log dividend-price ratio or log earnings-price ratio as the posited predictor variables. These measures are constructed using data on the monthly S&P stock price, dividends and earnings obtained from Robert J. Shiller's website at http://www.econ.yale.edu/∼shiller/data.htm. We use data for the period January 1871 to December 2014. The dividend-price ratio is calculated as the ratio of average dividends over the last year to the current stock price ($T = 1716$); the earnings-price ratio is calculated as the ratio of average dividends over the last ten years to the current stock price ($T = 1608$), cf. Campbell and Shiller (1988). Table 4 shows the values for the heteroskedasticity-robust IV_{comb} statistic of Breitung and Demetrescu (2015) and our S_B statistic, the latter implemented using BIC selection for the order of p in the fitted regression (12), starting from $p_{\max} = 12$ (roughly corresponding to $T^{1/3}$), with an appropriate degrees of freedom adjustment in the calculation of s_y^2. The P-values shown for IV_{comb} are for the $\chi^2(1)$ distribution, while those P-values shown for S_B are from the empirical bootstrap distribution based on $B = 9999$ bootstrap replications.

Table 4. Application to monthly U.S. Stock Index Returns, Jan. 1871 – Dec. 2014

	Predictor			
	Dividend-price ratio		Earnings-price ratio	
	IV_{comb}	S_B	IV_{comb}	S_B
Statistic	21.01	2.02	14.45	17.85
P-value	0.00	0.00	0.00	0.00

For the dividend-price ratio predictor we see that the IV_{comb} statistic rejects very strongly, thereby, when taken at face value, signalling substantial levels of predictability of S&P returns by the log dividend-price ratio.[6] However, any such conclusion of predictability is immediately thrown into serious question once we observe that S_B also rejects strongly, implying that such a predictive regression model is potentially spurious, or at the very least, under-specified by some latent persistent process. This pattern of results is similar when we consider the earnings-price ratio as an alternate potential predictor. If anything, when this predictor is employed the warning message provided via S_B is conveyed even more forcefully.[7]

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 unobserved persistent latent variables, themselves uncorrelated with the putative predictors under test, are present in the underlying data generation process. To guard against this possibility we have proposed a diagnostic test for such predictive regression 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. An empirical illustration using well-known U.S. stock market data highlighted the potential value of our procedure in practice.

References

Awad, A. (1981). Conditional central limit theorems for martingales and reversed martingales. Sankhya 43 A, 100–106.

Billingsley P. (1999). Convergence of Probability Measures. John Wiley & Sons, New York

Boswijk P., G. Cavaliere, A. Rahbek and A.M.R. Taylor (2015). Inference on co-integration parameters in heteroskedastic vector autoregressions. Forthcoming in Journal of Econometrics.

[6]Although not reported, the feasible Q statistic of Campbell and Yogo (2006) (implemented using BIC lag selection with $p_{\max} = 12$) indicated predictability at the 0.10-level.

[7]We also performed the IV_{comb} and S_B tests with the log dividend-price ratio and log earnings-price ratio *jointly* included in the predictive regression, as opposed to individually. P-values of 0.00 were found for these procedures also.

Breitung, J. and M. Demetrescu (2015). Instrumental variable and variable addition based inference in predictive regressions. Journal of Econometrics 187, 358–375.

Brown, B. (1971). Martingale central limit theorems. The Annals of Mathematical Statistics 42, 59–66.

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 Econometrics 81, 27–60.

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.

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

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.

Davidson, J. (1994). Stochastic Limit Theory. Oxford: Oxford University Press.

Davidson, R. and J.G. MacKinnon (2000). Bootstrap tests: how many bootstraps? Econometric Reviews 19, 55–68.

Deng, A. (2014). Understanding spurious regression in financial economics. Journal of Financial Econometrics 12, 122–150.

Denker, M. and M. Gordin (2003). On conditional central limit theorems for stationary processes. In Athreya, K., M. Majumdar, M. Puri and E. Waymire (eds.), Lecture Notes - Monograph Series Vol. 41, Probability, Statistics and Their Applications: Papers in Honor of Rabi Bhattacharya, Institute of Mathematical Statistics, Beachwood.

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.

Gonçalves, S. and L. Kilian (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. Journal of Econometrics 123, 89–120.

Gonçalves, S. and L. Kilian (2007). Asymptotic and bootstrap inference for $AR(\infty)$ processes with conditional heteroskedasticity. Econometric Reviews 26, 609–641.

Hansen, B.E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430.

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.

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.

Nyblom, J. (1989). Testing for the constancy of parameters over time. Journal of the American Statistical Association 84, 223–230.

Paye, B.S. and A. Timmermann (2006). Instability of return prediction models. Journal of Empirical Finance 13, 274–315.

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.

Appendix

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 $\mathring{y}_t := y_t - \bar{y}$, $\mathring{x}_t := x_t - \bar{x}_{-1}$ and $\Delta\mathring{x}_t := \Delta x_t - \overline{\Delta x}$, 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.* (2015),

$$T^{-1}\sum_{t=1}^T \epsilon_t\epsilon_t' \xrightarrow{p} H[\int_0^1 diag\{d_1^2(r), d_2^2(r), d_3^2(r)\}dr]H' \tag{18}$$

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

Fourth, we will also use the Orstein-Uhlenbeck convergence

$$T^{-1/2}\begin{bmatrix} x_{\lfloor T\cdot\rfloor} \\ z_{\lfloor T\cdot\rfloor} \end{bmatrix} \xrightarrow{w} \int_0^{\cdot} \begin{bmatrix} e^{-(\cdot-s)c_x}dM_{\eta x}(s) \\ e^{-(\cdot-s)c_z}dM_{\eta z}(s) \end{bmatrix} = \begin{bmatrix} M_{\eta x,c_x}(\cdot) \\ M_{\eta z,c_z}(\cdot) \end{bmatrix} =: M_{\eta c}(\cdot) \tag{19}$$

and the associated convergence to stochastic integrals

$$T^{-1}\sum_{t=1}^T \begin{bmatrix} x_{t-1} \\ z_{t-1} \end{bmatrix} [\epsilon_t', \Delta x_t, \Delta z_t] \xrightarrow{w} \int_0^1 M_{\eta c}(s)d[M_\eta(s)', M_{\eta c}(s)']. \tag{20}$$

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 α_y, α_x and α_z to zero, without loss of generality. First write t_u as

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

Then, we can write

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

and $T^{-2}\sum_{t=1}^T \mathring{x}_{t-1}^2 \xrightarrow{w} \int_0^1 \bar{M}_{\eta x,c_x}(r)^2$ using (19), (20) and the CMT. Also,

$$\begin{aligned} s_y^2 &= T^{-1}\sum_{t=1}^T \mathring{y}_t^2 - T^{-1}\frac{\{T^{-1}\sum_{t=1}^T \mathring{x}_{t-1}y_t\}^2}{T^{-2}\sum_{t=1}^T \mathring{x}_{t-1}^2} + o_p(1) = T^{-1}\sum_{t=1}^T y_t^2 - \bar{y}^2 + o_p(1) \\ &= T^{-1}\sum_{t=1}^T (g_x T^{-1}x_{t-1} + g_z T^{-1}z_{t-1} + \epsilon_{yt})^2 \\ &\quad - \left\{T^{-1}\sum_{t=1}^T (g_x T^{-1}x_{t-1} + g_z T^{-1}z_{t-1} + \epsilon_{yt})\right\}^2 + o_p(1) \\ &= T^{-1}\sum_{t=1}^T \epsilon_{yt}^2 + o_p(1) \xrightarrow{p} h_{31}^2\textstyle\int_0^1 d_1(r)^2 + h_{32}^2\int_0^1 d_2(r)^2 + h_{33}^2\int_0^1 d_3(r)^2 \end{aligned}$$

by (18). Consequently, by the CMT,

$$t_u \xrightarrow{w} \frac{g_x \int_0^1 \bar{M}_{\eta x,c_x}(r)^2 + g_z \int_0^1 \bar{M}_{\eta x,c_x}(r) M_{\eta z,c_z}(r) + \int_0^1 \bar{M}_{\eta x,c_x}(r) dM_{\eta y}(r)}{\sqrt{\left\{h_{31}^2 \int_0^1 d_1(r)^2 + h_{32}^2 \int_0^1 d_2(r)^2 + h_{33}^2 \int_0^1 d_3(r)^2\right\} \int_0^1 \bar{M}_{\eta x,c_x}(r)^2}}.$$

■

Proof of Theorem 2: It follows from the proof of Theorem 1 that

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

Also,

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

since $T^{-1}\sum_{t=1}^T \mathring{x}_{t-1}\epsilon_{xt} \xrightarrow{w} \int_0^1 \bar{M}_{\eta x,c_x}(r) dM_{\eta x}(r)$ using (19), (20) and the CMT. Now

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

giving

$$\begin{aligned}
s_x^2 &= T^{-1}\sum_{t=1}^T \{-(\hat{\rho}_x - \rho_x)\mathring{x}_{t-1} + \epsilon_{xt} - \bar{\epsilon}_x\}^2 + o_p(1) \\
&= (\hat{\rho}_x - \rho_x)^2 T^{-1}\sum_{t=1}^T \mathring{x}_{t-1}^2 + T^{-1}\sum_{t=1}^T (\epsilon_{xt} - \bar{\epsilon}_x)^2 \\
&\quad -2(\hat{\rho}_x - \rho_x) T^{-1}\sum_{t=1}^T \mathring{x}_{t-1}(\epsilon_{xt} - \bar{\epsilon}_x) + o_p(1) \\
&= T^{-1}\sum_{t=1}^T \epsilon_{xt}^2 + o_p(1) \xrightarrow{p} h_{11}^2 \int_0^1 d_1(r)^2
\end{aligned}$$

by (18), and

$$\begin{aligned}
s_{xy} &= T^{-1}\sum_{t=1}^T \hat{\epsilon}_{xt}\hat{\epsilon}_{yt} + o_p(1) \\
&= T^{-1}\sum_{t=1}^T \{-(\hat{\rho}_x - \rho_x)\mathring{x}_{t-1} + \epsilon_{xt} - \bar{\epsilon}_x\}\{\beta_x \mathring{x}_{t-1} + \beta_z \mathring{z}_{t-1} + (\epsilon_{yt} - \bar{\epsilon}_y) - \hat{\beta}_x \mathring{x}_{t-1}\} + o_p(1) \\
&= T^{-1}\sum_{t=1}^T \epsilon_{xt}\epsilon_{yt} + o_p(1) \xrightarrow{p} h_{11}h_{31}\int_0^1 d_1(r)^2
\end{aligned}$$

using (18).

So, using the limit of s_y^2 from Theorem 1, we find that

$$\begin{aligned} Q &= \frac{T\hat{\beta}_x - (s_{xy}/s_x^2)T(\hat{\rho}_x - \rho_x)}{\sqrt{s_y^2\{1-(s_{xy}^2/s_y^2 s_x^2)\}/T^{-2}\sum_{t=1}^T (x_{t-1}-\bar{x}_{-1})^2}} \\ &\xrightarrow{w} \frac{g_x\int_0^1 \bar{M}_{\eta x,c_x}(r)^2 + g_z\int_0^1 \bar{M}_{\eta x,c_x}(r)M_{\eta z,c_z}(r) + \int_0^1 \bar{M}_{\eta x,c_x}(r)dM_{\eta y}(r) - \frac{h_{31}}{h_{11}}\int_0^1 \bar{M}_{\eta x,c_x}(r)dM_{\eta x}(r)}{\sqrt{\begin{array}{c}\int_0^1 \bar{M}_{\eta x,c_x}(r)^2\left\{h_{31}^2\int_0^1 d_1(r)^2 + h_{32}^2\int_0^1 d_2(r)^2 + h_{33}^2\int_0^1 d_3(r)^2\right\} \\ \times\left[1 - h_{31}^2\int_0^1 d_1(r)^2\left\{h_{31}^2\int_0^1 d_1(r)^2 + h_{32}^2\int_0^1 d_2(r)^2 + h_{33}^2\int_0^1 d_3(r)^2\right\}^{-1}\right]\end{array}}} \\ &= \frac{g_x\int_0^1 \bar{M}_{\eta x,c_x}(r)^2 + g_z\int_0^1 \bar{M}_{\eta x,c_x}(r)M_{\eta z,c_z}(r) + \int_0^1 \bar{M}_{\eta x,c_x}(r)dM_{\eta y}(r) - \frac{h_{31}}{h_{11}}\int_0^1 \bar{M}_{\eta x,c_x}(r)dM_{\eta x}(r)}{\sqrt{\{h_{32}^2\int_0^1 d_2(r)^2 + h_{33}^2\int_0^1 d_3(r)^2\}\int_0^1 \bar{M}_{\eta x,c_x}(r)^2}} \end{aligned}$$

■

Proof of Theorem 3: We may set α_y, α_x and α_z to zero, and g_x to $-ch_{11}^{-1}h_{31}$, without loss of generality, since the $\hat{e}_t$ are invariant to these parameters. Let $y_t^x := y_t - h_{11}^{-1}h_{31}\Delta x_t$, $\mathring{y}_t^x := \mathring{y}_t - h_{11}^{-1}h_{33}\Delta\mathring{x}_t$ 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

$$\begin{aligned} T^{-1}\sum_{t=1}^T \mathring{x}_{t-1}y_t^x &= T^{-1}\sum_{t=1}^T \mathring{x}_{t-1}\epsilon_{yt}^x + g_z T^{-1}\sum_{t=1}^T \mathring{x}_{t-1}z_{t-1} \\ &\xrightarrow{w} \int_0^1 \bar{M}_{\eta x,c_x}(r)dB_\eta^*(r) + g_z\int_0^1 \bar{M}_{\eta x,c_x}(r)M_{\eta z,c_z}(r) \end{aligned} \tag{21}$$

using (19), (20) and the CMT.

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

$$T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\hat{e}_t = T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\mathring{y}_t - \left[\begin{array}{cc} T^{-3/2}\sum_{t=1}^{\lfloor Tr\rfloor}\mathring{x}_{t-1} & T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\Delta\mathring{x}_t \end{array}\right] N_T\hat{\boldsymbol{\beta}} \tag{22}$$

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

$$N_T\hat{\boldsymbol{\beta}} := \left[\begin{array}{cc} T^{-2}\sum_{t=1}^T \mathring{x}_{t-1}^2 & T^{-1}\sum_{t=1}^T \mathring{x}_{t-1}\Delta x_t \\ T^{-2}\sum_{t=1}^T \mathring{x}_{t-1}\Delta x_t & T^{-1}\sum_{t=1}^T (\Delta\mathring{x}_t)^2 \end{array}\right]^{-1}\left[\begin{array}{c} T^{-1}\sum_{t=1}^T \mathring{x}_{t-1}y_t \\ T^{-1}\sum_{t=1}^T \Delta\mathring{x}_t y_t \end{array}\right].$$

Before passing to the limit in (22), we focus on $N_T\hat{\boldsymbol{\beta}}$. It holds that

$$N_T\hat{\boldsymbol{\beta}} = \Delta_T^{-1}\left[\begin{array}{cc} T^{-1}\sum_{t=1}^T (\Delta\mathring{x}_t)^2 & -T^{-1}\sum_{t=1}^T \mathring{x}_{t-1}\Delta x_t \\ o_p(1) & T^{-2}\sum_{t=1}^T \mathring{x}_{t-1}^2 \end{array}\right]\left[\begin{array}{c} T^{-1}\sum_{t=1}^T \mathring{x}_{t-1}y_t \\ T^{-1}\sum_{t=1}^T \Delta\mathring{x}_t y_t \end{array}\right], \tag{23}$$

where $\Delta_T := T^{-3}\{\sum_{t=1}^T \mathring{x}_{t-1}^2\sum_{t=1}^T (\Delta\mathring{x}_t)^2 - (\sum_{t=1}^T \mathring{x}_{t-1}\Delta x_t)^2\} = T^{-3}\sum_{t=1}^T \mathring{x}_{t-1}^2\sum_{t=1}^T (\Delta\mathring{x}_t)^2 + o_p(T^{-3})$ because $\sum_{t=1}^T \mathring{x}_{t-1}\Delta x_t = O_p(T)$ by (19) and (20). Further, as also $\sum_{t=1}^T \mathring{x}_{t-1}y_t =$

$O_p(T)$ by the proof of Theorem 1, it holds that

$$\begin{aligned} N_T\hat{\boldsymbol{\beta}} &= \Delta_T^{-1}\left[\begin{array}{c} T^{-2}\{\sum_{t=1}^T \mathring{x}_{t-1}y_t \sum_{t=1}^T(\Delta\mathring{x}_t)^2 - \sum_{t=1}^T \mathring{x}_{t-1}\Delta x_t \sum_{t=1}^T \Delta\mathring{x}_t y_t\} \\ T^{-3}\sum_{t=1}^T \mathring{x}_{t-1}^2 \sum_{t=1}^T \Delta\mathring{x}_t y_t + o_p(1) \end{array}\right] \\ &= \Delta_T^{-1}\left[\begin{array}{c} T^{-2}\{\sum_{t=1}^T \mathring{x}_{t-1}y_t^x \sum_{t=1}^T(\Delta\mathring{x}_t)^2 - \sum_{t=1}^T \mathring{x}_{t-1}\Delta x_t \sum_{t=1}^T \Delta\mathring{x}_t y_t^x\} \\ T^{-3}\frac{h_{31}}{h_{11}}\sum_{t=1}^T \mathring{x}_{t-1}^2 \sum_{t=1}^T(\Delta\mathring{x}_t)^2 + T^{-3}\sum_{t=1}^T \mathring{x}_{t-1}^2 \sum_{t=1}^T \Delta\mathring{x}_t y_t^x + o_p(1) \end{array}\right] \\ &= \Delta_T^{-1}\left[\begin{array}{c} T^{-2}\sum_{t=1}^T \mathring{x}_{t-1}y_t^x \sum_{t=1}^T(\Delta\mathring{x}_t)^2 + o_p(1) \\ T^{-3}\frac{h_{31}}{h_{11}}\sum_{t=1}^T \mathring{x}_{t-1}^2 \sum_{t=1}^T(\Delta\mathring{x}_t)^2 + o_p(1) \end{array}\right] \end{aligned} \tag{24}$$

because $\sum_{t=1}^T \Delta\mathring{x}_t y_t^x = \sum_{t=1}^T \Delta x_t \epsilon_{yt}^x + g_z T^{-1}\sum_{t=1}^T \Delta x_t z_{t-1} - T^{-1}(x_T - x_1)\{\sum_{t=1}^T \epsilon_{yt}^x + g_z T^{-1}\sum_{t=1}^T z_{t-1}\}$ $= o_p(T)$ given that (i) $\sum_{t=1}^T \Delta x_t \epsilon_{yt}^x = \sum_{t=1}^T \epsilon_{xt}\epsilon_{yt}^x - cT^{-1}\sum_{t=1}^T x_{t-1}\epsilon_{yt}^x = o_p(T)$ using (18) and the convergence $T^{-1}\sum_{t=1}^T x_{t-1}\epsilon_{yt}^x \xrightarrow{w} \int_0^1 M_{\eta x,c_x}(s)dB_\eta^*(s)$ implied by (20), (ii) $T^{-1}\sum_{t=1}^T \Delta x_t z_{t-1}$ $\xrightarrow{w} \int_0^1 M_{\eta z,c_z}(r)dM_{\eta x,c_x}(r)$ as a consequence of (20), (iii) $T^{-1/2}(x_T - x_1) \xrightarrow{w} M_{\eta x,c_x}(1)$ by (19) and the CMT, (iv) $T^{-1/2}\sum_{t=1}^T \epsilon_{yt}^x \xrightarrow{w} B_\eta^*(1)$, and (v) $T^{-3/2}\sum_{t=1}^T z_{t-1} \xrightarrow{w} \int_0^1 M_{\eta z,c_z}(s)$ by (19) and the CMT. Finally,

$$N_T\hat{\boldsymbol{\beta}} = \left[\begin{array}{cc} (T^{-1}\sum_{t=1}^T \mathring{x}_{t-1}^2)^{-1}\sum_{t=1}^T \mathring{x}_{t-1}y_t^x & h_{31}^{-1}h_{33} \end{array}\right]' + o_p(1) \tag{25}$$

because $T^{-1}\sum_{t=1}^T(\Delta\mathring{x}_t)^2 = T^{-1}\sum_{t=1}^T \epsilon_{tx}^2 - 2c_x T^{-2}\sum_{t=1}^T \epsilon_{tx}x_{t-1} + T^{-3}c_x^2\sum_{t=1}^T x_{t-1}^2 - T^{-2}(x_T - x_1)^2 = T^{-1}\sum_{t=1}^T \epsilon_{tx}^2 + o_p(1) \xrightarrow{p} h_{11}^2\int_0^1 d_1^2(r)$ by (18), so $T^{-1}\sum_{t=1}^T(\Delta\mathring{x}_t)^2$ is bounded away from zero in P-probability.

Given (25), (22) simplifies to

$$T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\hat{e}_t = T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\mathring{y}_t^x - \frac{\sum_{t=1}^T \mathring{x}_{t-1}y_t^x}{T^{-1}\sum_{t=1}^T \mathring{x}_{t-1}^2}T^{-3/2}\sum_{t=1}^{\lfloor Tr\rfloor}\mathring{x}_{t-1} + \rho_T(r), \tag{26}$$

where

$$\begin{aligned} \sum_{t=1}^{\lfloor Tr\rfloor}\mathring{y}_t^x &= \sum_{t=1}^{\lfloor Tr\rfloor}\epsilon_{yt}^x + T^{-1}g_z\sum_{t=1}^{\lfloor Tr\rfloor}z_{t-1} - \frac{\lfloor Tr\rfloor - 1}{T}\{\sum_{t=1}^T \epsilon_{yt}^x + T^{-1}g_z\sum_{t=1}^T z_{t-1}\} \\ &\xrightarrow{w} B_\eta^*(r) - rB_\eta^*(1) + g_z(\int_0^r M_{\eta z,c_z}(s) - r\int_0^r M_{\eta z,c_z}) \end{aligned}$$

in the sense of weak convergence of measures on $\mathcal{D}$, and $\rho_T(r) = o_p(1)T^{-3/2}\sum_{t=1}^{\lfloor Tr\rfloor}\mathring{x}_{t-1} + o_p(1)T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\Delta\mathring{x}_t$ is such that

$$\sup_{r\in[0,1]}|\rho_T(r)| \le o_p(1)\sup_{r\in[0,1]}|T^{-3/2}\sum_{t=1}^{\lfloor Tr\rfloor}\mathring{x}_{t-1}| + o_p(1)T^{-1/2}\sup_{t=0,\ldots,T}|x_t| = o_p(1) \tag{27}$$

because $\sup_{r\in[0,1]}|T^{-3/2}\sum_{t=1}^{\lfloor Tr\rfloor}\mathring{x}_{t-1}| \xrightarrow{w} \sup_{r\in[0,1]}|\int_0^r \bar{M}_{\eta x,c_x}(s)|$ and $T^{-1/2}\sup_{t=0,\ldots,T}|x_t| \xrightarrow{w}$ $\sup_{r\in[0,1]}|M_{\eta x,c_x}(r)|$ by the CMT. Therefore, using also (21) and the CMT again,

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

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

$$\begin{aligned}\sum_{t=1}^{T}\hat{e}_t^2 &= \sum_{t=1}^{T}\mathring{y}_t^2-\left[\begin{array}{cc} T^{-1}\sum_{t=1}^{T}\mathring{x}_{t-1}y_t & \sum_{t=1}^{T}\Delta\mathring{x}_t y_t\end{array}\right]N_T\hat{\boldsymbol{\beta}} \qquad (28)\\ &= \sum_{t=1}^{T}\mathring{y}_t^2-h_{31}^{-1}h_{33}\sum_{t=1}^{T}\Delta\mathring{x}_t y_t-\sum_{t=1}^{T}\mathring{x}_{t-1}y_t(\sum_{t=1}^{T}\mathring{x}_{t-1}^2)^{-1}\sum_{t=1}^{T}\mathring{x}_{t-1}y_t^x+o_p(T)\\ &= \sum_{t=1}^{T}\mathring{y}_t^2-h_{31}^{-2}h_{33}^2\sum_{t=1}^{T}(\Delta\mathring{x}_t)^2-h_{31}^{-1}h_{33}\sum_{t=1}^{T}\Delta\mathring{x}_t y_t^x+o_p(T)\\ &= \sum_{t=1}^{T}(\mathring{y}_t^x)^2+h_{31}^{-1}h_{33}\sum_{t=1}^{T}y_t^x\Delta\mathring{x}_t+o_p(T)\\ &= \sum_{t=1}^{T}(\epsilon_{yt}^x)^2-2T^{-1}g_z\sum_{t=1}^{T}z_{t-1}\epsilon_{yt}+T^{-2}g_z^2\sum_{t=1}^{T}z_{t-1}^2+o_p(T)=\sum_{t=1}^{T}(\epsilon_{yt}^x)^2+o_p(T),\end{aligned}$$

where $T^{-1}\sum_{t=1}^{T}(\epsilon_{yt}^x)^2\xrightarrow{p}h_{32}^2\int_0^1 d_2(r)^2+h_{33}^2\int_0^1 d_3(r)^2$ by (18). Consequently,

$$s^2\xrightarrow{p}h_{32}^2\int_0^1 d_2(r)^2+h_{33}^2\int_0^1 d_3(r)^2, \qquad (29)$$

and by the CMT,

$$S\xrightarrow{w}\{h_{32}^2\int_0^1 d_2(r)^2+h_{33}^2\int_0^1 d_3(r)^2\}^{-1}\int_0^1\{F(r,c_x)+g_zG(r,c_x,c_z)\}^2dr.$$

■

Before proceeding to the proof of Theorem 4, 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.

Proof of Theorem 4: From Theorem 2 of Rubshtein (1996), by extending the argument to the multivariate case, it follows that $E(f(U_{\lfloor T\cdot\rfloor 2},U_{\lfloor T\cdot\rfloor 3})|\mathcal{X})\xrightarrow{a.s.}E\left(f\left(B_2,B_3\right)\right)$ for continuous bounded real f on $\mathcal{D}^2$. Then, by the bounded convergence theorem for conditional expectations,

$$E_xf(U_{\lfloor T\cdot\rfloor 2},U_{\lfloor T\cdot\rfloor 3})\xrightarrow{a.s.}Ef\left(B_2,B_3\right) \qquad (30)$$

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

$$E_x f(U'_{\lfloor T\cdot\rfloor}) \xrightarrow{w} E(f(B')|B_1) \tag{31}$$

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

Next, we note that U_{tb}, given the data, is a Gaussian process with independent increments, mean zero and variance function $V_T(r) := Var^*(U_{\lfloor Tr\rfloor b}) = T^{-1}\sum_{t=1}^{\lfloor Tr\rfloor} e_t e_t' \xrightarrow{p} rI_3$ ($r \in [0,1]$), 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

$$E^* f(U'_{\lfloor T\cdot\rfloor b}) \xrightarrow{p} Ef(B^{\dagger\prime}) \tag{32}$$

for continuous bounded real f on $\mathcal{D}^3$. Additionally, $[U'_{\lfloor T\cdot\rfloor}, U'_{\lfloor T\cdot\rfloor b}]' \xrightarrow{w} [B', B^{\dagger\prime}]'$ 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 $\mathcal{D}^6$,

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

here we have used the result that conditioning on x, y, z and $U_{\lfloor T\cdot\rfloor}$ are equivalent. In particular, for f that do not depend on $U_{\lfloor T\cdot\rfloor 1}, U_{\lfloor T\cdot\rfloor 2}$, restricted to $\mathcal{D}^4$, the bootstrap counterpart of (31) is obtained:

$$E^* f(U_{\lfloor T\cdot\rfloor 1}, U'_{\lfloor T\cdot\rfloor b}) \xrightarrow{w} E\{f(B_1, B^{\dagger\prime})|B\} = E\{f(B_1, B^{\dagger\prime})|B_1\}, \tag{33}$$

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

To see that (31) and (33) are joint, it is sufficient to apply the Cramer-Wald device to obtain

$$aE_x f(U'_{\lfloor T\cdot\rfloor}) + bE^* g(U_{\lfloor T\cdot\rfloor 1}, U'_{\lfloor T\cdot\rfloor b}) \xrightarrow{w} E(af(B') + bg(B_1, B^{\dagger\prime})|B_1) \tag{34}$$

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 $[B', B^{\dagger\prime}]'$ has a.s. continuous sample paths, we can consider a probability space where $[U_{\lfloor T\cdot\rfloor}, U'_{\lfloor T\cdot\rfloor b}]' \to [B', B^{\dagger\prime}]'$ a.s. On this probability space, by Corollary 4.4 of Crimaldi and Pratelli (2005), (31) and (33) hold in probability instead of weakly, and hence, (34) holds in probability. Since the distribution of the involved conditional expectations only depends on $[U'_{\lfloor T\cdot\rfloor}, U'_{\lfloor T\cdot\rfloor b}]'$ and $[B', B^{\dagger\prime}]'$, it follows that on general probability spaces (34) holds weakly. ■

Proof of Theorem 5: Introduce $\tilde{\epsilon}_{it} := d_t e_{it}$, $\tilde{U}_{ti} := T^{-1/2}\sum_{s=1}^{t}\tilde{\epsilon}_{is}$, $\tilde{M}_i(\cdot) := \int_0^{\cdot} d_i(s)dB_i(s)$ ($i = 1, 2, 3$), $\tilde{U}_t := [\tilde{U}_{t1}, \tilde{U}_{t2}, \tilde{U}_{t3}]'$, $\tilde{M} := [\tilde{M}_1, \tilde{M}_2, \tilde{M}_3]'$. Given that ϵ_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

$$\left(\tilde{U}_{\lfloor T\cdot\rfloor}, \sum_{t=1}^{T}\tilde{U}_{t-1,1}[\Delta\tilde{U}_{t2}, \Delta\tilde{U}_{t3}]\right) \xrightarrow{w_x} \left(\tilde{M}, \int_0^1 \tilde{M}_1(s)d[\tilde{M}_2(s), \tilde{M}_3(s)]\right)\Bigg| B_1 \tag{35}$$

jointly with

$$\left(U_{\lfloor T\cdot\rfloor 1}, \tilde{U}_{\lfloor T\cdot\rfloor b}, \sum_{t=1}^{T} \tilde{U}_{t-1,1}\Delta\tilde{U}_{tb}\right) \xrightarrow{w^*} \left(B_1, \tilde{B}_m^{\dagger}, \int_0^1 \tilde{M}_1(s)d\tilde{B}_m^{\dagger}(s)\right)\Bigg| B_1. \tag{36}$$

We shall prove Theorem 5 in this way.

Notice first that, given the data, $\tilde{U}_{\lfloor T\cdot\rfloor b}$ is a Gaussian process with independent increments, mean zero and variance function $Var^*(\tilde{U}_{\lfloor Tr\rfloor b}) = T^{-1}\sum_{t=1}^{\lfloor Tr\rfloor}\tilde{e}_{Tt}^2$. 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 $\tilde{U}_{\lfloor T\cdot\rfloor b}$ given the data (and thus, given $U_{\lfloor T\cdot\rfloor}$) converges weakly in probability to $\tilde{B}_m^{\dagger}$:

$$E^*g(\tilde{U}_{\lfloor T\cdot\rfloor b}) \xrightarrow{p} Eg(\tilde{B}_m^{\dagger}) \tag{37}$$

for all bounded continuous real g on $\mathcal{D}$, where $\tilde{B}_m^{\dagger}$ is a Gaussian process with independent increments, zero mean and variance function $\int_0^{\cdot} m^2(s)ds$. On the other hand, since $U_{\lfloor T\cdot\rfloor} \xrightarrow{w} B$ by the MFCLT of Brown (1971), and since $\mathcal{D}^3\times\mathcal{D}$ is separable, it follows that $[U'_{\lfloor T\cdot\rfloor}, \tilde{U}_{\lfloor T\cdot\rfloor b}]' \xrightarrow{w} [B', \tilde{B}_m^{\dagger}]'$ on $\mathcal{D}^3\times\mathcal{D}$, with B and $\tilde{B}_m^{\dagger}$ 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 $[B', \tilde{B}_m^{\dagger}]'$'s sample paths, we may assume in the remainder of the proof that $[U'_{\lfloor T\cdot\rfloor}, \tilde{U}_{\lfloor T\cdot\rfloor b}]'$ and $[B', \tilde{B}_m^{\dagger}]'$ are defined on the same probability space (say $\mathbb{S}$), and

$$[U'_{\lfloor T\cdot\rfloor}, \tilde{U}_{\lfloor T\cdot\rfloor b}]' \to [B', \tilde{B}_m^{\dagger}]' \text{ a.s.} \tag{38}$$

By using (38) and the distributional properties of $[U'_{\lfloor T\cdot\rfloor}, \tilde{U}_{\lfloor T\cdot\rfloor b}]'$ (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 (35)-(36) 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 $\mathbb{S}$,

$$E_x\phi\left(\tilde{U}'_{\lfloor T\cdot\rfloor}, \sum_{t=1}^{T}\tilde{U}_{t-1,1}[\Delta\tilde{U}_{t2}, \Delta\tilde{U}_{t3}]\right) \xrightarrow{p} E\left[\phi\left(\tilde{M}', \int_0^1 \tilde{M}_1(s)d[\tilde{M}_2(s), \tilde{M}_3(s)]\right)\Bigg| B_1\right] \tag{39}$$

and

$$E^*\psi\left(U_{\lfloor T\cdot\rfloor 1}, \tilde{U}_{\lfloor T\cdot\rfloor, b}, \sum_{t=1}^{T}\tilde{U}_{t-1,1}\Delta\tilde{U}_{tb}\right) \xrightarrow{p} E\left[\psi\left(B_1, \tilde{B}_m^{\dagger}, \int_0^1 \tilde{M}_1(s)d\tilde{B}_m^{\dagger}(s)\right)\Bigg| B_1\right] \tag{40}$$

for every bounded and continuous real ϕ and ψ 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 $E(\cdot|U_{\lfloor T\cdot\rfloor 1})$ and $E(\cdot|U_{\lfloor T\cdot\rfloor})$. In view of (30) and (37), on $\mathbb{S}$ we can still invoke

$$E_x f(U_{\lfloor T\cdot\rfloor 2}, U_{\lfloor T\cdot\rfloor 3}) \xrightarrow{w} Ef\left(B_2, B_3\right) \text{ and } E^*g(\tilde{U}_{\lfloor T\cdot\rfloor b}) \xrightarrow{w} Eg(\tilde{B}_m^{\dagger})$$

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 $[U'_{\lfloor T\cdot\rfloor}, \tilde{U}_{\lfloor T\cdot\rfloor b}]'$

and $[B', \tilde{B}_m^{\dagger}]'$. Moreover, in view also of (38), by Corollary 4.4 of Crimaldi and Pratelli (2005), it holds on $\mathbb{S}$ that

$$E_x h(U'_{\lfloor T\cdot\rfloor}) \xrightarrow{p} E\{h(B')|B_1\} \text{ and } E^* g(U_{\lfloor T\cdot\rfloor 1}, \tilde{U}_{\lfloor T\cdot\rfloor b}) \xrightarrow{p} E\{g(B_1, \tilde{B}_m^{\dagger})|B_1\} \tag{41}$$

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

It is well known that (39)-(40) cannot be put in the form of (41) 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) $\tilde{U}_{\lfloor T\cdot\rfloor j}$ by $\xi_{\delta j}(U_{\lfloor T\cdot\rfloor j})$ $(j = 1, 2, 3)$, where $\xi_{\delta j} : \mathcal{D} \to \mathcal{D}$ 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 $\delta_j : [0, 1] \to \mathbb{R}$. Then, using (41) and integration by parts, it follows that

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

and

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

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

$$E^* n(U_{\lfloor T\cdot\rfloor 1}, \tilde{U}_{\lfloor T\cdot\rfloor 1}, \tilde{U}_{\lfloor T\cdot\rfloor b}) \xrightarrow{p} E\{n(B_1, \tilde{M}_1, \tilde{B}_m^{\dagger})|B_1\}.$$

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

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

with

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

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

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

and

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

for ϕ and ψ as in (39)-(40). To complete the proof, it remains to be shown that, as $L \to \infty$, ζ_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,

$$\max_{r\in[0,1]} |\tilde{U}_{\lfloor Tr\rfloor j} - \xi_{\delta j}(U_{\lfloor T\cdot\rfloor j})(r)| \leq \max_{r\in[0,1]} \left| \frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor rT\rfloor} \{d_j(\frac{t}{T}) - \delta_j(\frac{t}{T})\}\Delta U_{tj} \right| + \frac{1}{2} \max_{r\in[0,1]} |R_T(r)|, \quad (42)$$

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

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

because $\{\max_{t=1,\ldots,T} |U_{tj}|\}|x \to \max_{[0,1]} |B_j|$ (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

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

Since smooth functions are dense in $L_2[0,1]$, this limit can be made as close to 1 as desired by choosing δ_1 according to λ. 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

$$\begin{aligned} & P_x \left(\max_{r\in[0,1]} \left| \frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor rT\rfloor} \{d_j(\frac{t}{T}) - \delta_j(\frac{t}{T})\}\Delta U_{tj} \right| \geq \lambda \right) \quad (43) \\ \leq \;& \frac{1}{\lambda^2} E_x \left(\frac{1}{T^{1/2}} \sum_{t=1}^{T} \{d_j(\frac{t}{T}) - \delta_j(\frac{t}{T})\}\Delta U_{tj} \right)^2 = \frac{1}{\lambda^2 T} \sum_{t=1}^{T} \{d_j(\frac{t}{T}) - \delta_j(\frac{t}{T})\}^2 E_x[(\Delta U_{tj})^2] \end{aligned}$$

and from Markov's inequality that

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

which can be made as small as desired by the choice of δ_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 δ_j, as in step (a.1).

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

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

Here, first, $\tilde{U}_{\lfloor T\cdot\rfloor 1}\xrightarrow{p}\tilde{M}_1$ can be established on $\mathbb{S}$ by using the approximation of $\tilde{U}_{\lfloor T\cdot\rfloor 1}$ with $\xi_{\delta 1}(U_{\lfloor T\cdot\rfloor 1})$ 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 $T\to\infty$. Indeed, $E_x\gamma_{Tij}=\Gamma_{Tij,K}^{\leq}+\Gamma_{Tij,K}^{>}$ for every $K>0$, where

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

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

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

Further, using Doob's sub-martingale inequality,

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

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

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

On the other hand, it also holds that

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

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

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

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

as $T\to\infty$, 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. ■

We next discuss some implications of Theorem 5 for Orstein-Uhlenbeck limits and stochastic integrals involving them. With $s_{x,0}=\alpha_x=0$, the standard evaluation

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

holds for almost all x, by the ergodic theorem. As $\sum_{i=1}^{\lfloor Tr\rfloor}e^{c_x\frac{i}{T}}\epsilon_{xi}=h_{11}\sum_{i=1}^{\lfloor Tr\rfloor}e^{c_x\frac{i}{T}}d_1(\frac{i}{T})e_{1i}$, by applying Theorem 5 with $e^{c_x(\cdot)}d_1(\cdot)$ in place of $d_1(\cdot)$, it follows that

$$
T^{-1/2}x_{\lfloor T\cdot\rfloor}\xrightarrow{w_x}h_{11}e^{-c_x(\cdot)}\int_0^{\cdot}e^{c_xs}d_1(s)dB_1(s)\Big|B_1=M_{\eta x,c_x}(\cdot)|B_1,
$$

and similarly, $T^{-1/2}z_{\lfloor T\cdot\rfloor} \overset{w_x}{\to} M_{\eta z,c_z}(\cdot)|B_1$, jointly with the convergence in Theorem 5, by the argument for that theorem.

Regarding stochastic integrals, for $\tilde{\epsilon}_{it}$ $(i = 2, 3)$ introduced in the proof of Theorem 5, we find by partial summation that

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

where the following converge by the CMT, Theorem 5 and the discussion in the previous paragraph: $T^{-1}s_{x,T}\sum_{t=1}^{T}\tilde{\epsilon}_{it} \overset{w_x}{\to} M_{\eta x,c_x}(1)\tilde{M}_i(1)|B_1$, $T^{-1}\sum_{t=1}^{T}\epsilon_{xt}\sum_{s=1}^{t-1}\tilde{\epsilon}_{is} \overset{w_x}{\to} h_{11}\int_0^1[d\tilde{M}_1(s)]\tilde{M}_i(s)|B_1$, $T^{-2}\sum_{t=1}^{T}s_{x,t-1}\sum_{s=1}^{t-1}\tilde{\epsilon}_{is} \overset{w_x}{\to} h_{11}\int_0^1\tilde{M}_1(s)\tilde{M}_i(s)ds|B_1$ jointly. Moreover, $T^{-1}\sum_{t=1}^{T}\epsilon_{xt}\tilde{\epsilon}_{it} = o_p^x(1)$ by the conditional Chebyshev inequality, as

$$T^{-1}Var_x(\sum_{t=1}^{T}\epsilon_{xt}\tilde{\epsilon}_{it}) \le KT^{-1}\sum_{t=1}^{T}e_{1t}^2 E_x e_{it}^2 \to KE(e_{1t}^2 e_{it}^2) \text{ a.s.} \tag{44}$$

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

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}s_{x,t-1}\tilde{\epsilon}_{it} &\overset{w_x}{\to} \left(M_{\eta x,c_x}(1)\tilde{M}_i(1) - h_{11}\int_0^1[d\tilde{M}_1(s)]\tilde{M}_i(s) + c_x h_{11}\int_0^1\tilde{M}_1(s)\tilde{M}_i(s)ds\right)\Bigg| B_1 \\ &= \int_0^1\tilde{M}_i(s)dM_{\eta x,c_x}(s)\Bigg| B_1 \end{aligned}$$

jointly with the convergence in Theorem 5 and its implications. By continuity again, as $T^{-2}\sum_{t=1}^{T}s_{x,t-1}z_{t-1} \overset{w_x}{\to} \int_0^1 M_{\eta x,c_x}(s)\,M_{\eta z,c_z}(s)\,ds|B_1$ and $T^{-3/2}\sum_{t=1}^{T-1}s_{x,t} \overset{w_x}{\to} M_{\eta x,c_x}(1)|B_1$, it follows for $\mathring{s}_{x,t} := s_{x,t} - T^{-1}\sum_{i=1}^{T-1}s_{x,i}$ and $\epsilon_{yt}^x := \epsilon_{yt} - h_{31}d_{1t}e_{1t}$ that

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}\mathring{s}_{x,t-1}y_t^x &= T^{-1}\sum_{t=1}^{T}\mathring{s}_{x,t-1}(\epsilon_{yt}^x + T^{-1}g_z z_{t-1}) \\ &\overset{w_x}{\to} \left\{\int_0^1\bar{M}_{\eta x,c_x}(s)\,dB_\eta^*(s) + g_z\int_0^1\bar{M}_{\eta x,c_x}(s)\,M_{\eta z,c_z}(s)\,ds\right\}\Bigg| B_1, \end{aligned} \tag{45}$$

if $g_x = 0$, where B_η^* is defined in Theorem 3.

Proof of Theorem 6: We again set α_y, α_x, α_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 ξ_T,

$$\xi_T \overset{p}{\to} K = const \qquad \text{implies that} \qquad \xi_T \overset{w_x}{\to} K \tag{46}$$

because $\xi_T \overset{p}{\to} K$ implies, for bounded continuous f, that $E_x f(\xi_T) \overset{p}{\to} f(K)$. This follows from the 'in probability' bounded convergence theorem for conditional expectations.

From relations (26)-(27), with $\xi_T = \sup_{r\in[0,1]}|\rho_T(r)|$, it follows that

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

uniformly in r. Here, from Theorem 5, the convergence $T^{-1/2}z_{\lfloor T\cdot\rfloor} \overset{w_x}{\to} M_{\eta z,c_z}(\cdot)|B_1$ and the CMT,

$$\begin{aligned}\sum_{t=1}^{\lfloor T\cdot\rfloor}\mathring{y}_t^x &= \sum_{t=1}^{\lfloor T\cdot\rfloor}\epsilon_{yt}^x + T^{-1}g_z\sum_{t=1}^{\lfloor T\cdot\rfloor}z_{t-1} - \frac{\lfloor T\cdot\rfloor - 1}{T}\{\sum_{t=1}^{T}\epsilon_{yt}^x + T^{-1}g_z\sum_{t=1}^{T}z_{t-1}\} \\ &\overset{w_x}{\to} \{B_\eta^*(\cdot) - (\cdot)B_\eta^*(1) + g_z(\int_0^\cdot M_{\eta z,c_z}(s)\,ds - (\cdot)\int_0^1 M_{\eta z,c_z}(s)\,ds)\}\Big| B_1,\end{aligned}$$

so using also (45), the convergence $T^{-1/2}x_{\lfloor T\cdot\rfloor} \overset{w_x}{\to} M_{\eta x,c_x}(\cdot)|B_1$ and the CMT, we have that

$$T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\hat{e}_t \overset{w_x}{\to} \{F(\cdot,c_x) + g_z G(\cdot,c_x,c_z)\}|B_1.$$

Next, (29) and (46) with $\xi_T = s_y^2$ imply that $s_y^2 \overset{w_x}{\to} h_{32}^2\int_0^1 d_2(r)^2 + h_{33}^2\int_0^1 d_3(r)^2$. Consequently, by the CMT,

$$S \overset{w_x}{\to} (\{h_{32}^2\textstyle\int_0^1 d_2(r)^2 + h_{33}^2\int_0^1 d_3(r)^2\}^{-1}\int_0^1\{F(r,c_x) + g_z G(r,c_x,c_z)\}^2 dr)|B_1. \tag{47}$$

We proceed to part (b). The bootstrap process $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}y_t^*$ is of the form of $\tilde{U}_{\lfloor T\cdot\rfloor b}$ of Theorem 5, 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 $T^{-1}\sum_{t=1}^{\lfloor Tr\rfloor}\hat{e}_t^2 \overset{p}{\to} h_{32}^2\int_0^r d_2^2(s)ds + h_{33}^2\int_0^r d_3^2(s)ds = \int_0^r m^2(s)ds$ 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 $Var(B_\eta^{\dagger *}(r)) = \int_0^r m^2(s)\,ds$, from Theorem 5 and its discussion it follows that

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

jointly with $T^{-1/2}x_{\lfloor T\cdot\rfloor} \overset{w^*}{\to} M_{\eta x,c_x}|B_1$ and (47).

Next,

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

where by the CMT, the following converge jointly, and jointly with (47): $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}(y_t^* - \bar{y}^*) \overset{w^*}{\to} \{B_\eta^{\dagger *}(\cdot) - (\cdot)B_\eta^{\dagger *}(1)\}|B_1$, $T^{-3/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\mathring{x}_{t-1} \overset{w^*}{\to} \int_0^\cdot \bar{M}_{\eta x,c_x}(s)ds|B_1$, $T^{-1}\sum_{t=1}^{T}\mathring{x}_{t-1}y_t^* \overset{w^*}{\to} \int_0^1 \bar{M}_{\eta x,c_x}(s)dB_\eta^{\dagger *}(s)|B_1$ analogously to (45), $T^{-2}\sum_{t=1}^{T}\mathring{x}_{t-1}^2 \overset{w^*}{\to} \int_0^1 \bar{M}_{\eta x,c_x}^2(s)ds|B_1$, and since the two limit processes in $\mathcal{D}$ are continuous,

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

in $\mathcal{D}$, jointly with (47). Moreover, using the previous convergence results we have that,

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

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 $s_y^{*2} \xrightarrow{w^*} h_{32}^2\int d_2^2 + h_{33}^2\int d_3^2$ and, by the CMT, that

$$S^* \xrightarrow{w^*} \left(\{h_{32}^2\int d_2^2 + h_{33}^2\int d_3^2\}^{-1}\int F^{\dagger}(r, c_x)^2 dr\right)\Big| B_1$$

jointly with (47). ■

Proof of Corollary 1: The asymptotic validity of the bootstrap rests on the result that, as $T \to \infty$, 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 6, 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 f, where S_∞ is implicitly defined by (16). 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 Φ denotes the cumulative process of S_∞ conditional on B_1 (i.e., $\Phi(z) := P(S_\infty \leq z|B_1)$, all z), then $[P_x(S \leq \cdot), P^*(S^* \leq \cdot)]' \xrightarrow{w} [\Phi, \Phi]'$ in $\mathcal{D} \times \mathcal{D}$. As the distribution of S_∞ conditional on B_1 is atomless a.s., and so Φ is continuous a.s., the latter convergence holds also in $\mathcal{D}^2$ and implies that $\sup_{x \in \mathbb{R}} |P_x(S \leq x) - P^*(S^* \leq x)| = o_p(1)$. Therefore, if Φ_T denotes the cumulative process of S conditional on x (i.e., $\Phi_T(z) := P_x(S \leq z)$, all z), then $P^*(S^* \leq S) = \Phi_T(S) + o_p(1)$.

Further, define the quantile transformation using the right-continuous version of the generalised inverse. As the quantile transformation is continuous in the Skorokhod metric, it holds that $(\Phi_T, \Phi_T^{-1}) \xrightarrow{w} (\Phi, \Phi^{-1})$ in $\mathcal{D}^2$. For $\theta \in [0, 1]$,

$$\begin{aligned}
P_x(\Phi_T(S) \geq \theta) &= P_x(S \geq \Phi_T^{-1}(\theta)) = 1 - P_x(S < \Phi_T^{-1}(\theta)) \\
&= 1 - \Phi_T(\Phi_T^{-1}(\theta)-) \xrightarrow{w} 1 - \Phi(\Phi^{-1}(\theta)) = \theta
\end{aligned}$$

using the continuity of Φ, and the same holds in probability as the limit is a constant. By the Bounded convergence theorem, integration over x yields $P(\Phi_T(S) \geq \theta) \to \theta$ for $\theta \in [0, 1]$. Therefore, $\Phi_T(S) \xrightarrow{w} U[0, 1]$. Since $P^*(S^* \leq S) = \Phi_T(S) + o_p(1)$, it also holds that $P^*(S^* \leq S) \xrightarrow{w} U[0, 1]$.

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

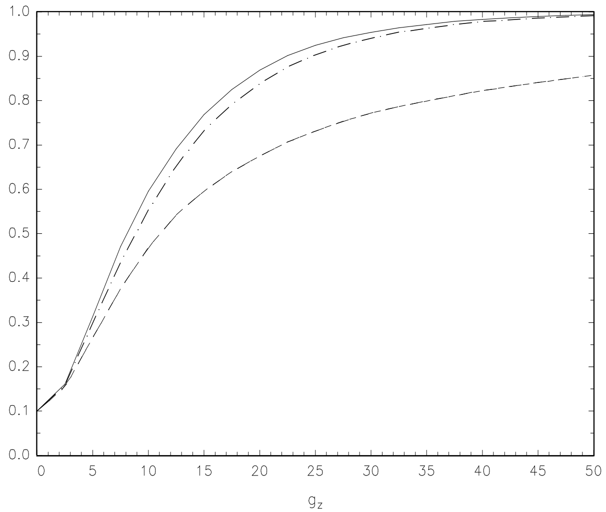
		$c_x = 0$						$c_x = 5$						$c_x = 10$					
		$\tau = 0.3$			$\tau = 0.7$			$\tau = 0.3$			$\tau = 0.7$			$\tau = 0.3$			$\tau = 0.7$		
σ_1	σ_3	S_B	IV_{comb}	IV_{comb}^{pre}	S_B	IV_{comb}	IV_{comb}^{pre}	S_B	IV_{comb}	IV_{comb}^{pre}	S_B	IV_{comb}	IV_{comb}^{pre}	S_B	IV_{comb}	IV_{comb}^{pre}	S_B	IV_{comb}	IV_{comb}^{pre}
1	1	0.098	0.109	0.094	0.098	0.109	0.094	0.103	0.104	0.089	0.103	0.104	0.089	0.102	0.107	0.093	0.102	0.107	0.093
	4	0.101	0.108	0.094	0.101	0.113	0.089	0.106	0.107	0.094	0.105	0.113	0.093	0.105	0.108	0.096	0.107	0.111	0.093
	$\frac{1}{4}$	0.102	0.112	0.081	0.098	0.106	0.087	0.104	0.107	0.079	0.099	0.106	0.089	0.104	0.105	0.081	0.102	0.105	0.091
4	1	0.100	0.110	0.093	0.102	0.112	0.096	0.103	0.106	0.091	0.104	0.112	0.093	0.104	0.109	0.094	0.104	0.116	0.098
	4	0.099	0.109	0.099	0.102	0.118	0.104	0.107	0.110	0.098	0.107	0.120	0.106	0.106	0.114	0.103	0.109	0.123	0.109
	$\frac{1}{4}$	0.101	0.109	0.062	0.099	0.100	0.073	0.104	0.104	0.062	0.102	0.101	0.071	0.106	0.102	0.064	0.102	0.105	0.078
$\frac{1}{4}$	1	0.102	0.112	0.093	0.099	0.111	0.094	0.102	0.108	0.091	0.105	0.107	0.090	0.104	0.109	0.092	0.110	0.107	0.091
	4	0.103	0.107	0.079	0.103	0.110	0.074	0.102	0.101	0.082	0.108	0.107	0.077	0.104	0.101	0.081	0.108	0.107	0.081
	$\frac{1}{4}$	0.103	0.115	0.099	0.098	0.108	0.091	0.105	0.113	0.095	0.101	0.109	0.093	0.106	0.113	0.097	0.101	0.110	0.097

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

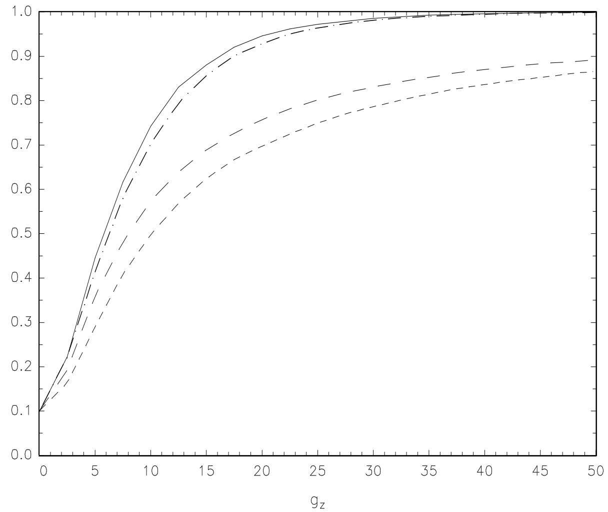
			$c_x = c_z = 0$						$c_x = c_z = 5$						$c_x = c_z = 10$					
			$\tau = 0.3$			$\tau = 0.7$			$\tau = 0.3$			$\tau = 0.7$			$\tau = 0.3$			$\tau = 0.7$		
σ_1	σ_2	σ_3	S_B	IV_{comb}	IV_{comb}^{pre}	S_B	IV_{comb}	IV_{comb}^{pre}	S_B	IV_{comb}	IV_{comb}^{pre}	S_B	IV_{comb}	IV_{comb}^{pre}	S_B	IV_{comb}	IV_{comb}^{pre}	S_B	IV_{comb}	IV_{comb}^{pre}
1	1	1	0.910	0.714	0.049	0.910	0.714	0.049	0.742	0.382	0.075	0.742	0.382	0.075	0.568	0.245	0.087	0.568	0.245	0.087
		4	0.478	0.434	0.192	0.585	0.504	0.158	0.252	0.162	0.112	0.308	0.199	0.115	0.174	0.130	0.105	0.198	0.146	0.106
		$\frac{1}{4}$	0.970	0.770	0.018	0.944	0.742	0.034	0.880	0.491	0.043	0.828	0.423	0.054	0.754	0.342	0.063	0.688	0.279	0.068
	4	1	0.997	0.845	0.002	0.977	0.767	0.013	0.985	0.638	0.007	0.919	0.539	0.029	0.960	0.479	0.015	0.842	0.404	0.043
		4	0.905	0.736	0.052	0.815	0.624	0.075	0.761	0.401	0.075	0.612	0.348	0.087	0.585	0.242	0.085	0.462	0.230	0.088
		$\frac{1}{4}$	0.999	0.857	0.001	0.987	0.780	0.008	0.995	0.674	0.003	0.947	0.569	0.021	0.986	0.534	0.006	0.895	0.439	0.032
	$\frac{1}{4}$	1	0.656	0.553	0.129	0.864	0.703	0.068	0.469	0.253	0.106	0.661	0.348	0.088	0.340	0.182	0.100	0.481	0.218	0.095
		4	0.245	0.263	0.179	0.534	0.487	0.167	0.153	0.128	0.105	0.251	0.189	0.126	0.131	0.117	0.102	0.168	0.139	0.107
		$\frac{1}{4}$	0.817	0.646	0.066	0.904	0.739	0.053	0.641	0.355	0.084	0.754	0.392	0.070	0.482	0.248	0.083	0.588	0.249	0.080
4	1	1	0.907	0.721	0.054	0.912	0.687	0.041	0.739	0.387	0.079	0.745	0.389	0.074	0.569	0.240	0.087	0.576	0.256	0.086
		4	0.464	0.416	0.198	0.602	0.380	0.112	0.254	0.159	0.116	0.317	0.170	0.107	0.175	0.127	0.106	0.204	0.140	0.110
		$\frac{1}{4}$	0.971	0.801	0.019	0.942	0.769	0.036	0.885	0.563	0.053	0.823	0.524	0.073	0.751	0.419	0.082	0.680	0.380	0.091
	4	1	0.996	0.858	0.003	0.968	0.755	0.022	0.982	0.644	0.010	0.907	0.583	0.048	0.956	0.479	0.019	0.828	0.463	0.069
		4	0.896	0.737	0.063	0.781	0.553	0.096	0.754	0.385	0.080	0.577	0.325	0.111	0.579	0.229	0.086	0.432	0.226	0.119
		$\frac{1}{4}$	0.999	0.872	0.001	0.978	0.783	0.016	0.993	0.691	0.004	0.940	0.641	0.036	0.983	0.550	0.009	0.882	0.539	0.059
	$\frac{1}{4}$	1	0.679	0.545	0.104	0.886	0.662	0.040	0.487	0.245	0.096	0.688	0.326	0.065	0.351	0.170	0.093	0.505	0.207	0.080
		4	0.253	0.248	0.168	0.576	0.355	0.103	0.158	0.127	0.107	0.279	0.154	0.105	0.135	0.119	0.104	0.178	0.133	0.108
		$\frac{1}{4}$	0.826	0.689	0.062	0.919	0.757	0.036	0.660	0.412	0.080	0.771	0.474	0.062	0.494	0.296	0.078	0.601	0.313	0.074
$\frac{1}{4}$	1	1	0.909	0.711	0.044	0.914	0.723	0.046	0.744	0.375	0.067	0.733	0.367	0.078	0.573	0.255	0.085	0.567	0.235	0.084
		4	0.494	0.504	0.186	0.584	0.569	0.173	0.255	0.201	0.121	0.291	0.258	0.146	0.174	0.145	0.102	0.190	0.178	0.122
		$\frac{1}{4}$	0.975	0.738	0.011	0.943	0.739	0.033	0.874	0.416	0.036	0.828	0.383	0.051	0.755	0.287	0.052	0.687	0.246	0.062
	4	1	0.996	0.848	0.001	0.979	0.767	0.010	0.988	0.623	0.004	0.913	0.498	0.026	0.965	0.473	0.010	0.838	0.345	0.034
		4	0.920	0.774	0.036	0.824	0.665	0.069	0.791	0.471	0.058	0.614	0.386	0.085	0.606	0.305	0.079	0.466	0.254	0.082
		$\frac{1}{4}$	0.999	0.855	0.000	0.989	0.773	0.006	0.996	0.638	0.001	0.946	0.507	0.016	0.988	0.493	0.003	0.894	0.355	0.023
	$\frac{1}{4}$	1	0.603	0.574	0.181	0.855	0.721	0.078	0.444	0.283	0.132	0.649	0.357	0.102	0.323	0.214	0.129	0.468	0.226	0.104
		4	0.214	0.326	0.223	0.515	0.554	0.201	0.150	0.154	0.114	0.235	0.248	0.161	0.129	0.131	0.102	0.161	0.172	0.127
		$\frac{1}{4}$	0.785	0.613	0.093	0.897	0.738	0.059	0.596	0.320	0.107	0.750	0.368	0.073	0.448	0.238	0.114	0.585	0.234	0.083

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

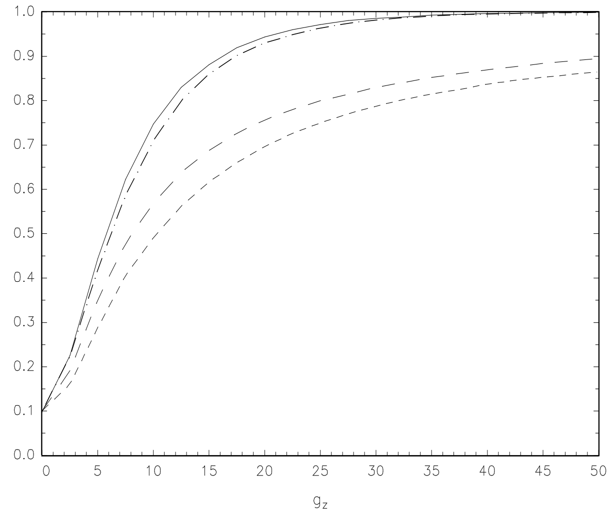
			$c_x = c_z = 0$						$c_x = c_z = 5$						$c_x = c_z = 10$					
			$\tau = 0.3$			$\tau = 0.7$			$\tau = 0.3$			$\tau = 0.7$			$\tau = 0.3$			$\tau = 0.7$		
σ_1	σ_2	σ_3	S_B	IV_{comb}	IV_{comb}^{pre}	S_B	IV_{comb}	IV_{comb}^{pre}	S_B	IV_{comb}	IV_{comb}^{pre}	S_B	IV_{comb}	IV_{comb}^{pre}	S_B	IV_{comb}	IV_{comb}^{pre}	S_B	IV_{comb}	IV_{comb}^{pre}
1	1	1	0.987	0.808	0.009	0.987	0.808	0.009	0.944	0.545	0.023	0.944	0.545	0.023	0.866	0.393	0.040	0.866	0.393	0.040
		4	0.761	0.622	0.120	0.848	0.677	0.078	0.527	0.276	0.110	0.607	0.345	0.107	0.356	0.176	0.102	0.414	0.222	0.107
		$\frac{1}{4}$	0.996	0.835	0.003	0.992	0.820	0.005	0.981	0.621	0.008	0.968	0.577	0.015	0.949	0.481	0.020	0.924	0.427	0.025
	4	1	1.000	0.863	0.000	0.996	0.815	0.002	0.997	0.689	0.001	0.976	0.636	0.011	0.992	0.557	0.004	0.944	0.520	0.021
		4	0.984	0.819	0.010	0.956	0.748	0.023	0.946	0.566	0.023	0.857	0.518	0.048	0.871	0.392	0.044	0.740	0.375	0.065
		$\frac{1}{4}$	1.000	0.866	0.000	0.998	0.820	0.002	0.999	0.701	0.001	0.983	0.647	0.009	0.996	0.575	0.002	0.958	0.534	0.016
	$\frac{1}{4}$	1	0.886	0.710	0.054	0.973	0.807	0.015	0.777	0.420	0.063	0.908	0.523	0.035	0.639	0.297	0.076	0.809	0.364	0.052
		4	0.465	0.441	0.197	0.796	0.670	0.098	0.293	0.186	0.121	0.530	0.314	0.121	0.210	0.143	0.108	0.339	0.206	0.119
		$\frac{1}{4}$	0.951	0.768	0.022	0.983	0.820	0.009	0.882	0.531	0.041	0.948	0.561	0.021	0.793	0.406	0.051	0.882	0.401	0.034
4	1	1	0.988	0.812	0.008	0.987	0.786	0.008	0.942	0.552	0.026	0.946	0.549	0.022	0.860	0.391	0.046	0.874	0.402	0.040
		4	0.759	0.609	0.123	0.861	0.572	0.051	0.520	0.257	0.108	0.625	0.271	0.079	0.353	0.162	0.100	0.431	0.183	0.094
		$\frac{1}{4}$	0.996	0.849	0.003	0.992	0.825	0.006	0.981	0.656	0.011	0.970	0.638	0.016	0.948	0.528	0.022	0.925	0.513	0.032
	4	1	0.999	0.874	0.001	0.994	0.790	0.004	0.996	0.690	0.003	0.977	0.649	0.012	0.990	0.544	0.005	0.944	0.546	0.026
		4	0.982	0.825	0.012	0.943	0.696	0.032	0.940	0.553	0.029	0.845	0.490	0.061	0.866	0.366	0.045	0.726	0.364	0.084
		$\frac{1}{4}$	1.000	0.877	0.000	0.996	0.800	0.003	0.998	0.703	0.001	0.984	0.669	0.009	0.994	0.567	0.003	0.961	0.573	0.019
	$\frac{1}{4}$	1	0.891	0.708	0.047	0.978	0.776	0.009	0.784	0.410	0.055	0.922	0.501	0.021	0.649	0.276	0.063	0.826	0.337	0.038
		4	0.487	0.418	0.167	0.829	0.545	0.052	0.297	0.170	0.112	0.575	0.227	0.076	0.218	0.136	0.105	0.370	0.158	0.093
		$\frac{1}{4}$	0.957	0.801	0.020	0.985	0.826	0.008	0.894	0.584	0.036	0.952	0.621	0.018	0.797	0.464	0.049	0.889	0.476	0.032
$\frac{1}{4}$	1	1	0.987	0.805	0.009	0.987	0.810	0.008	0.947	0.545	0.018	0.941	0.535	0.024	0.869	0.409	0.040	0.857	0.382	0.044
		4	0.773	0.681	0.113	0.847	0.729	0.084	0.536	0.349	0.122	0.589	0.435	0.146	0.356	0.239	0.124	0.400	0.292	0.146
		$\frac{1}{4}$	0.997	0.816	0.001	0.992	0.817	0.005	0.980	0.573	0.007	0.969	0.549	0.012	0.949	0.437	0.017	0.924	0.390	0.024
	4	1	0.999	0.872	0.000	0.997	0.821	0.002	0.998	0.691	0.001	0.975	0.618	0.011	0.993	0.573	0.003	0.939	0.483	0.019
		4	0.985	0.841	0.008	0.958	0.778	0.022	0.955	0.616	0.018	0.852	0.551	0.053	0.884	0.462	0.035	0.736	0.414	0.068
		$\frac{1}{4}$	1.000	0.874	0.000	0.998	0.825	0.002	0.999	0.697	0.001	0.983	0.622	0.008	0.997	0.584	0.001	0.954	0.489	0.014
	$\frac{1}{4}$	1	0.849	0.721	0.090	0.970	0.815	0.018	0.749	0.455	0.093	0.901	0.529	0.043	0.623	0.351	0.108	0.798	0.368	0.063
		4	0.407	0.530	0.271	0.790	0.730	0.114	0.285	0.253	0.153	0.498	0.421	0.177	0.204	0.199	0.139	0.320	0.281	0.166
		$\frac{1}{4}$	0.946	0.743	0.026	0.982	0.823	0.011	0.851	0.487	0.058	0.944	0.542	0.023	0.761	0.386	0.073	0.878	0.378	0.038



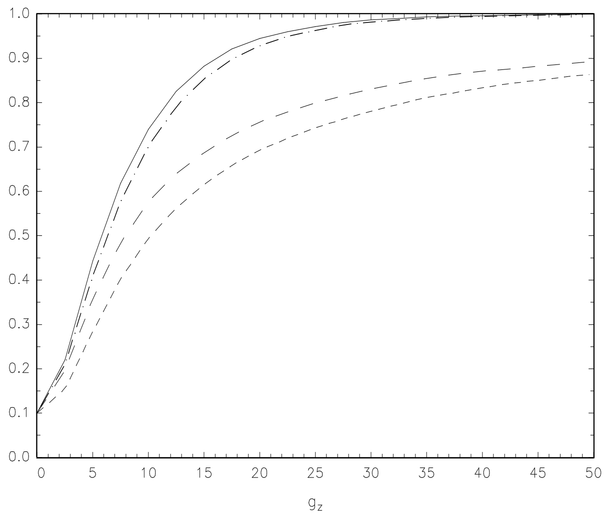
(a) $\sigma_{xy} = 0$, $\sigma_{zy} = 0$



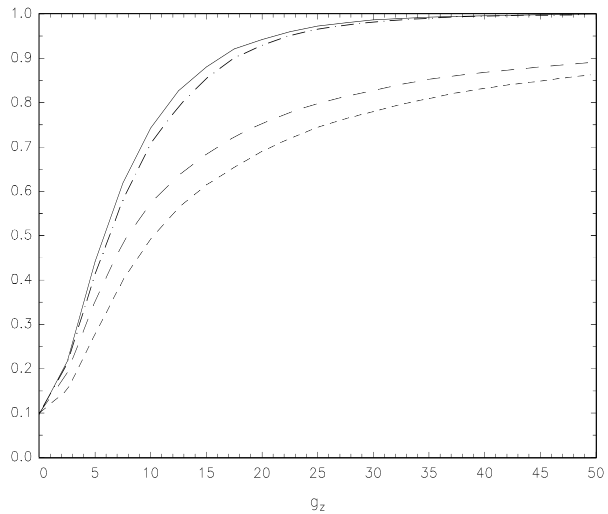
(b) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.35$



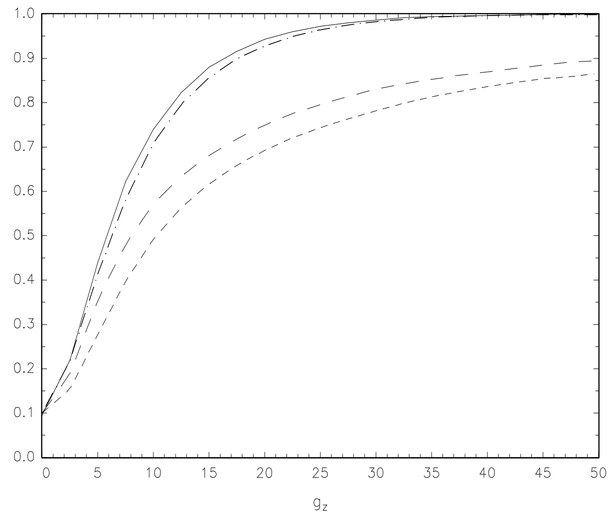
(c) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$



(d) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0$

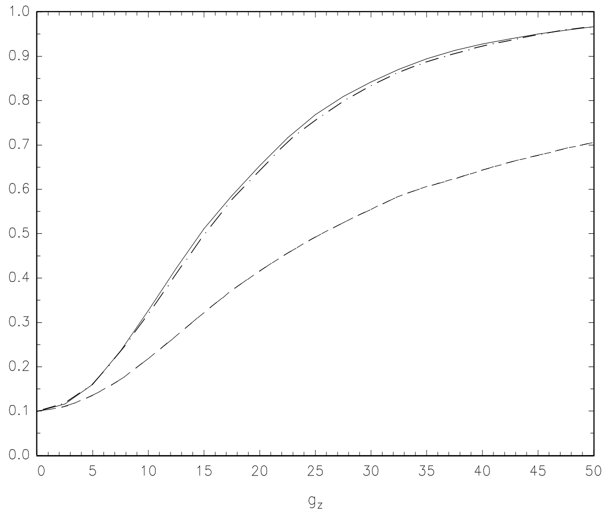


(e) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0.35$

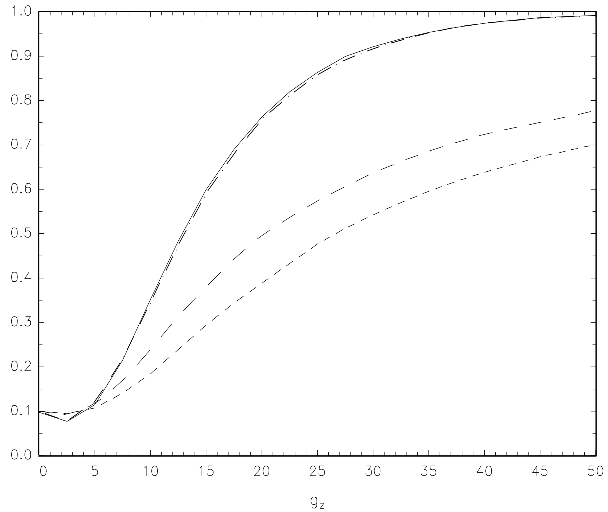


(f) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$

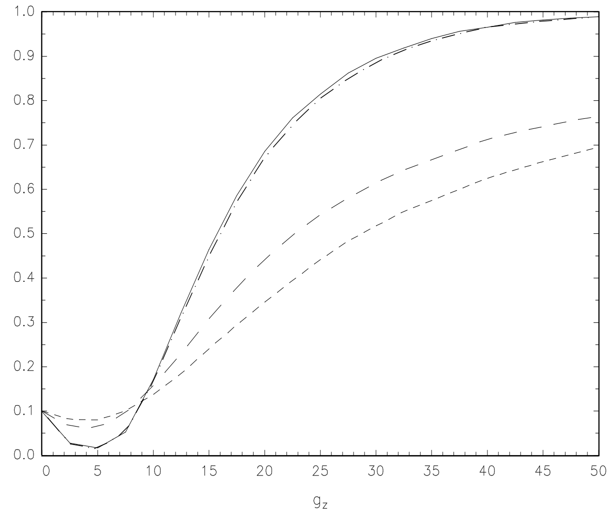
Figure 1. Asymptotic rejection frequencies of S, S_B (power) and t_u, Q (size): $g_x = 0$, $c_x = c_z = 0$;
S: $- \cdot -$, S_B: —— , t_u: - - - , Q: $- -$



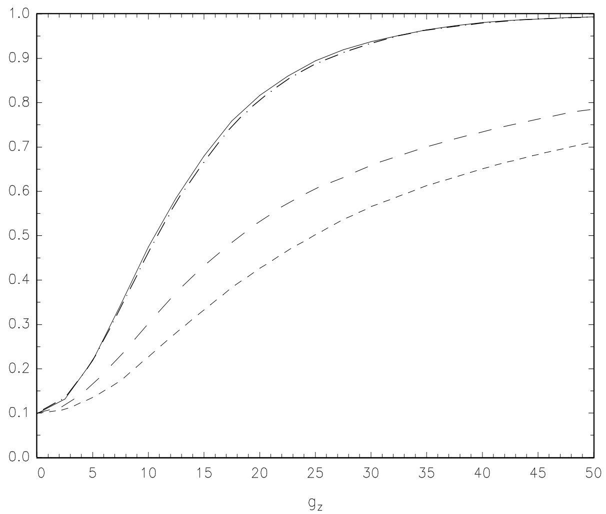
(a) $\sigma_{xy} = 0$, $\sigma_{zy} = 0$



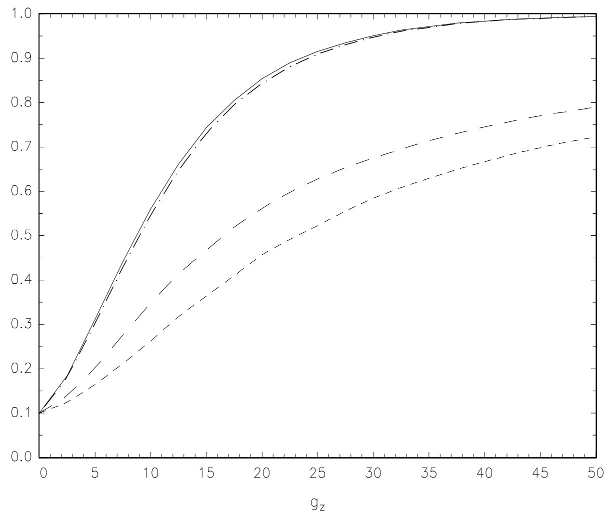
(b) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.35$



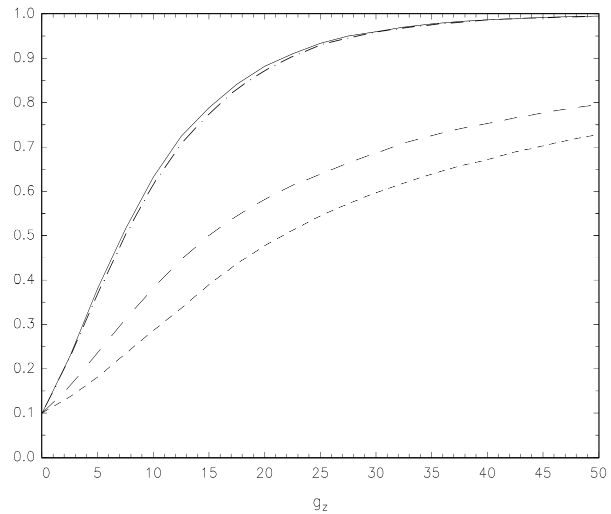
(c) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$



(d) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0$

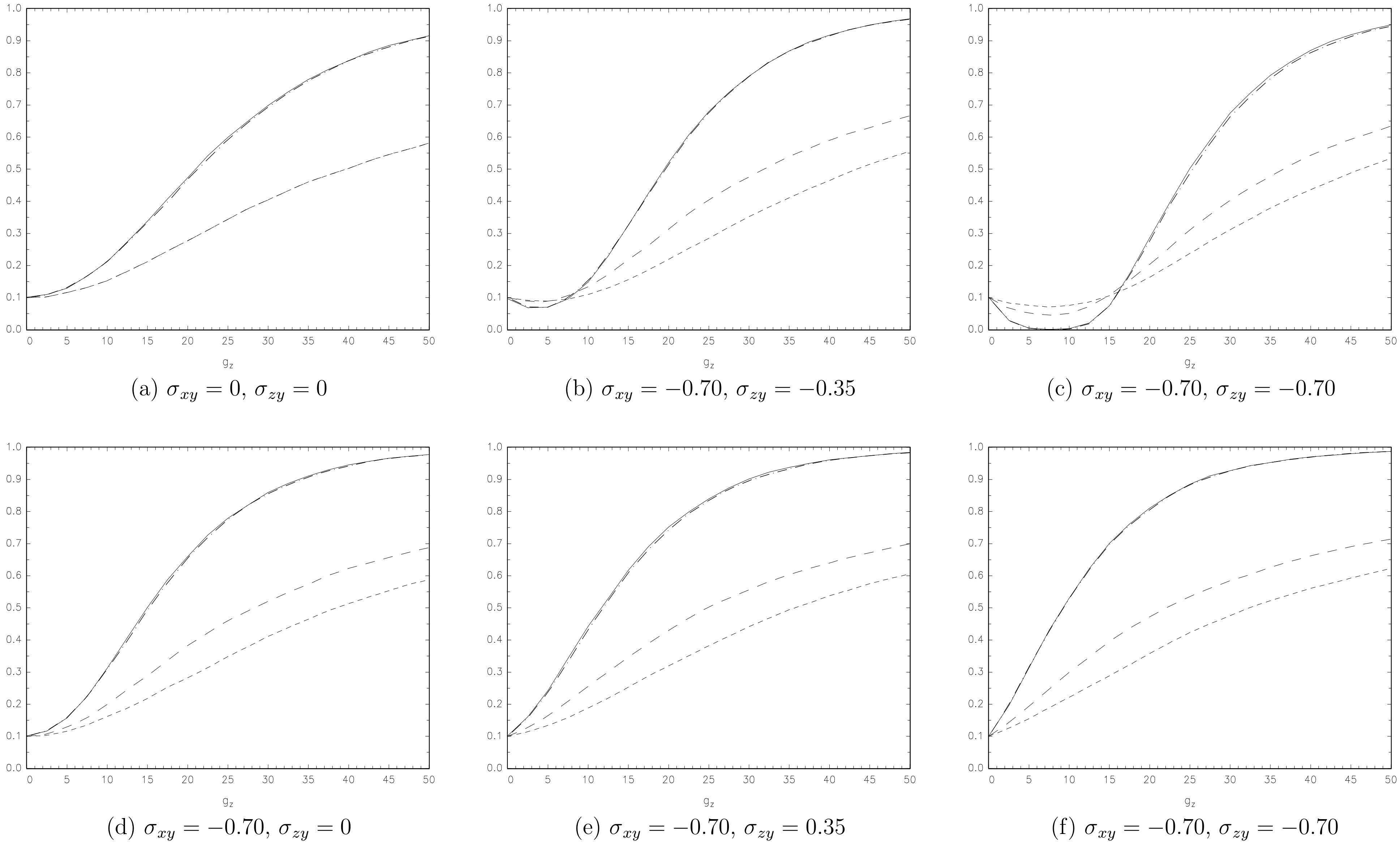


(e) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0.35$



(f) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$

Figure 2. Asymptotic rejection frequencies of S, S_B (power) and t_u, Q (size): $g_x = 0$, $c_x = c_z = 5$;
S: $- \cdot -$, S_B: —— , t_u: - - - , Q: $- -$



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

(b) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.35$

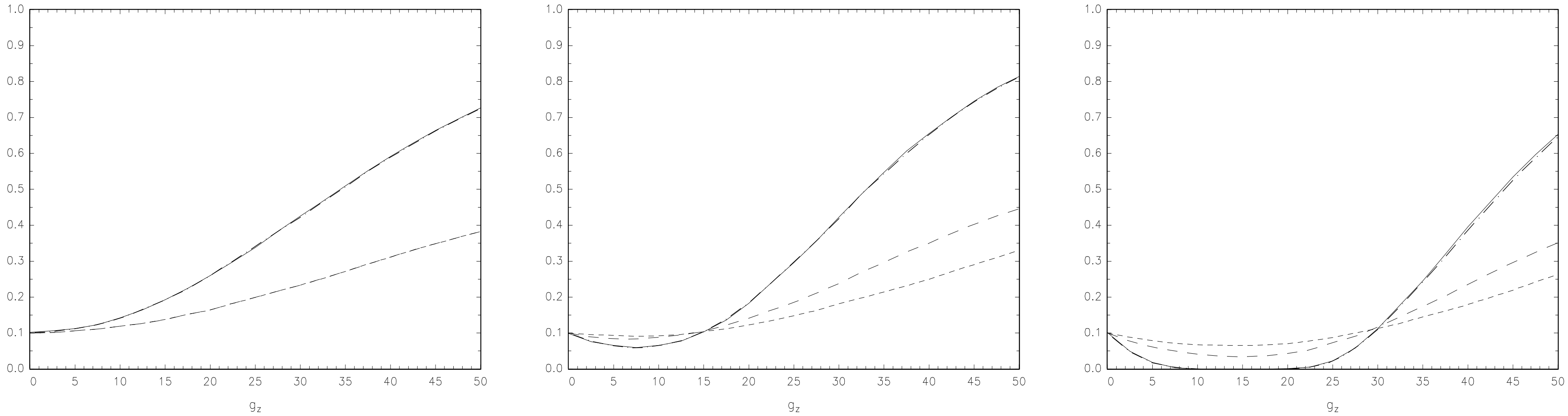
(c) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$

(d) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0$

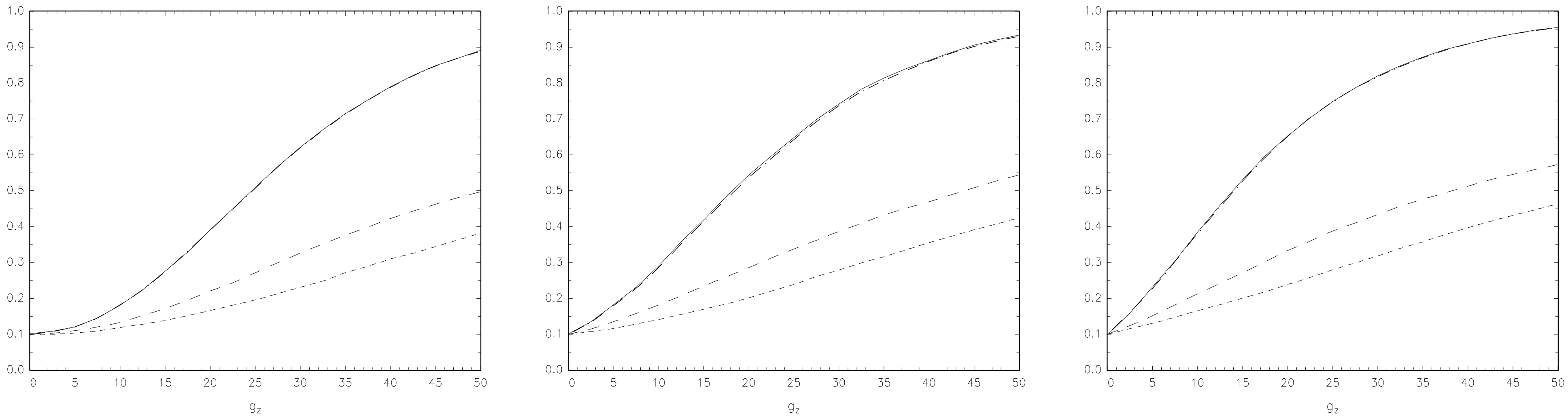
(e) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0.35$

(f) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$

Figure 3. 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: $- -$

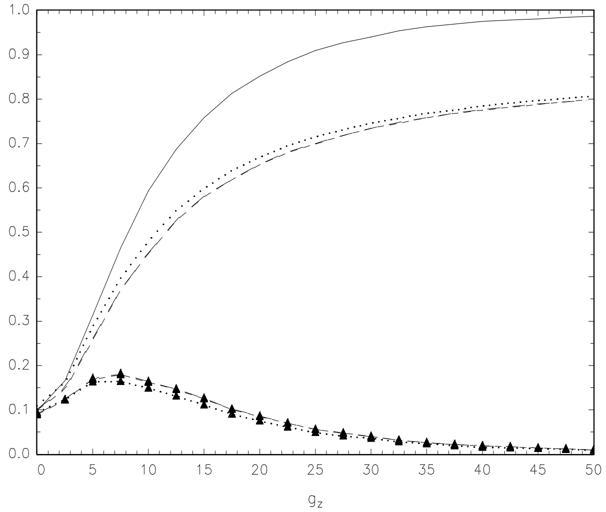


(a) $\sigma_{xy} = 0$, $\sigma_{zy} = 0$ (b) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.35$ (c) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$

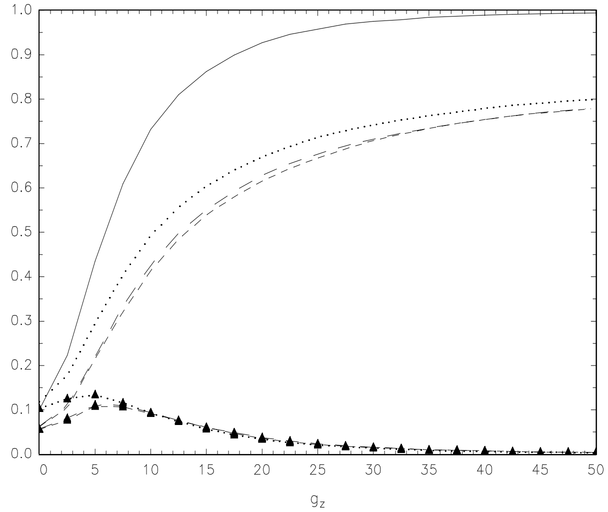


(d) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0$ (e) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0.35$ (f) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$

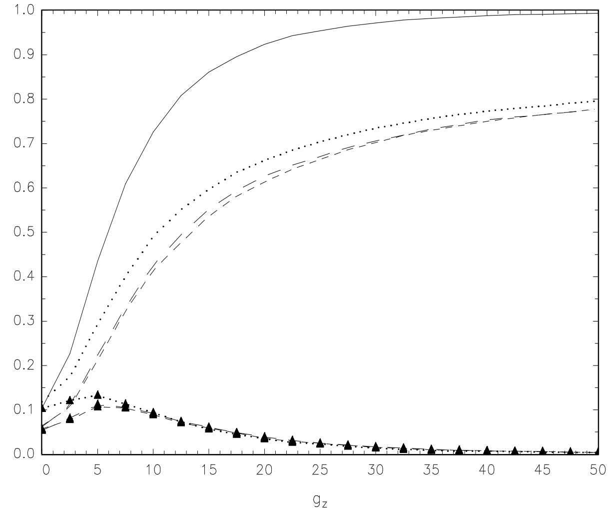
Figure 4. Asymptotic rejection frequencies of S, S_B (power) and t_u, Q (size): $g_x = 0$, $c_x = c_z = 20$;
S: $- \cdot -$, S_B: —— , t_u: - - - , Q: $- -$



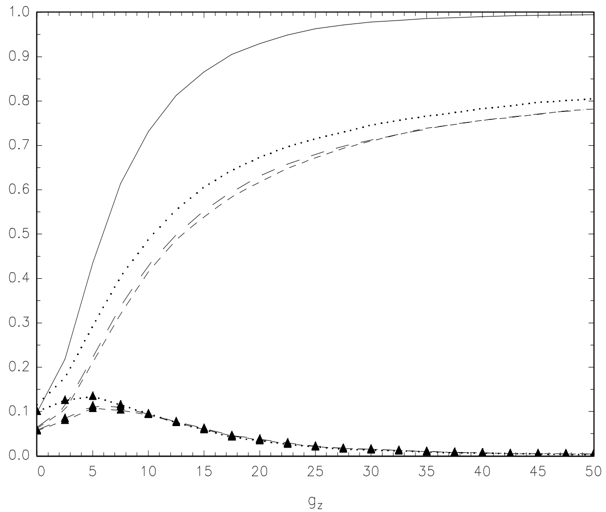
(a) $\sigma_{xy} = 0$, $\sigma_{zy} = 0$



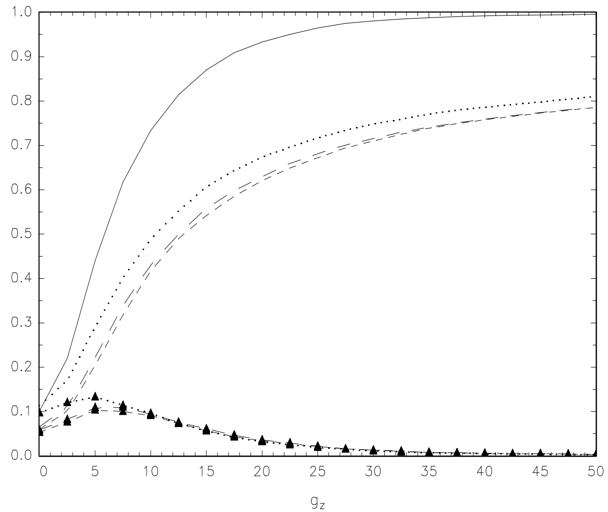
(b) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.35$



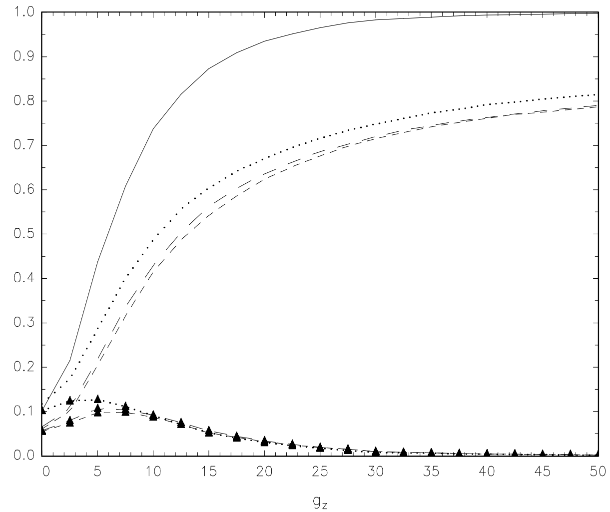
(c) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$



(d) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0$

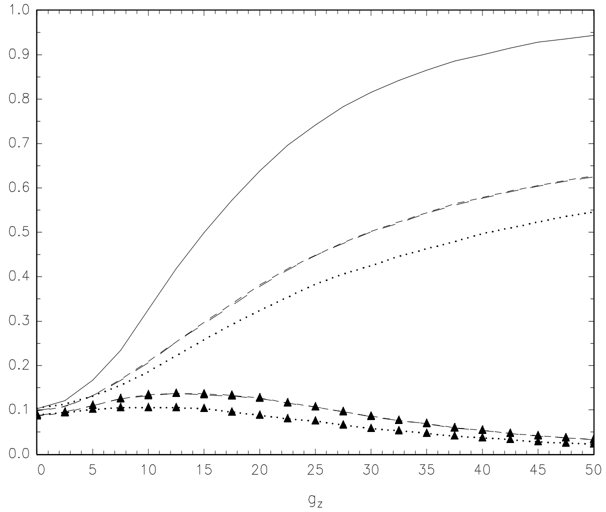


(e) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0.35$

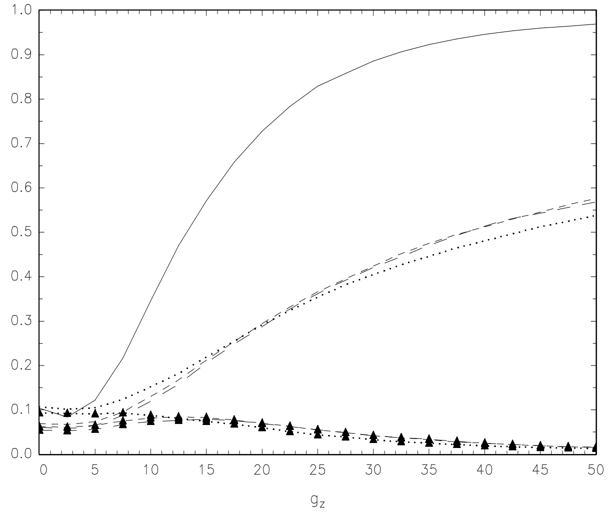


(f) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$

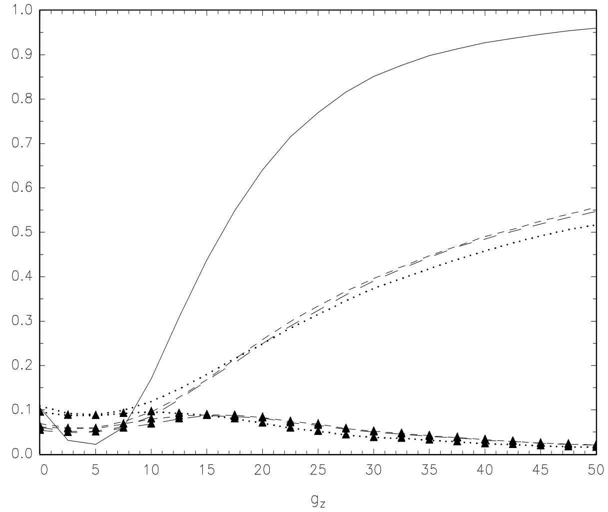
Figure 5. Finite sample rejection frequencies of S_B (power) and t_u, Q, IV_{comb}, t_u^{pre}, Q^{pre}, IV_{comb}^{pre} (size): $T = 200$, $g_x = 0$, $c_x = c_z = 0$;
S_B: —— , t_u: - - - , Q: – – , IV_{comb}: $\cdots$, t_u^{pre}: - ▲ - , Q^{pre}: – ▲ – , IV_{comb}^{pre}: $\cdots$▲$\cdots$



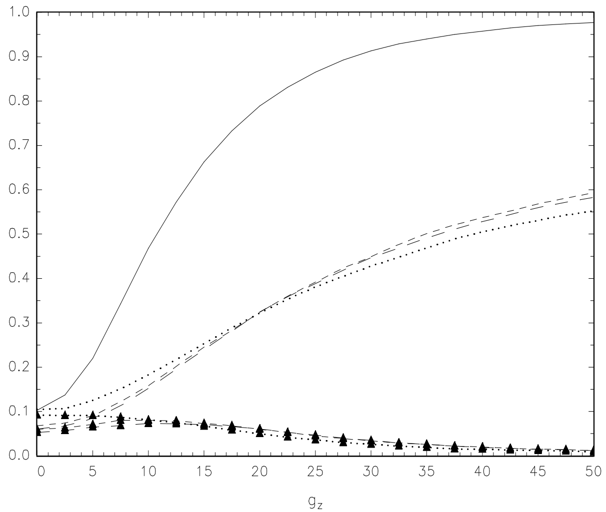
(a) $\sigma_{xy} = 0$, $\sigma_{zy} = 0$



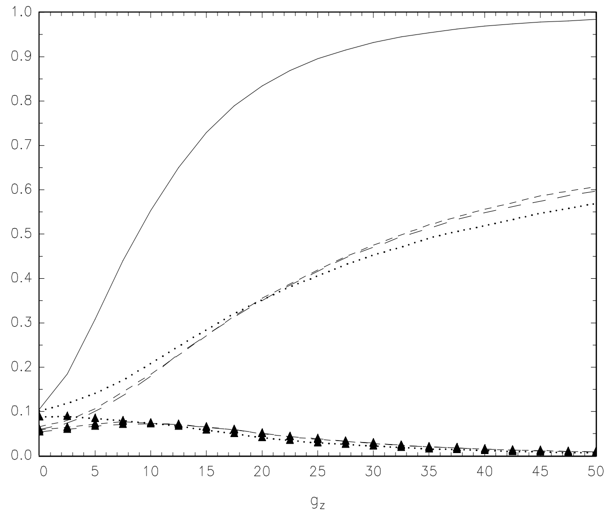
(b) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.35$



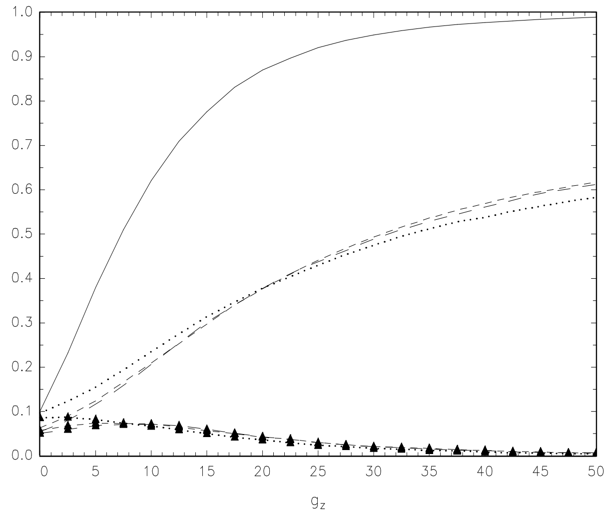
(c) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$



(d) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0$

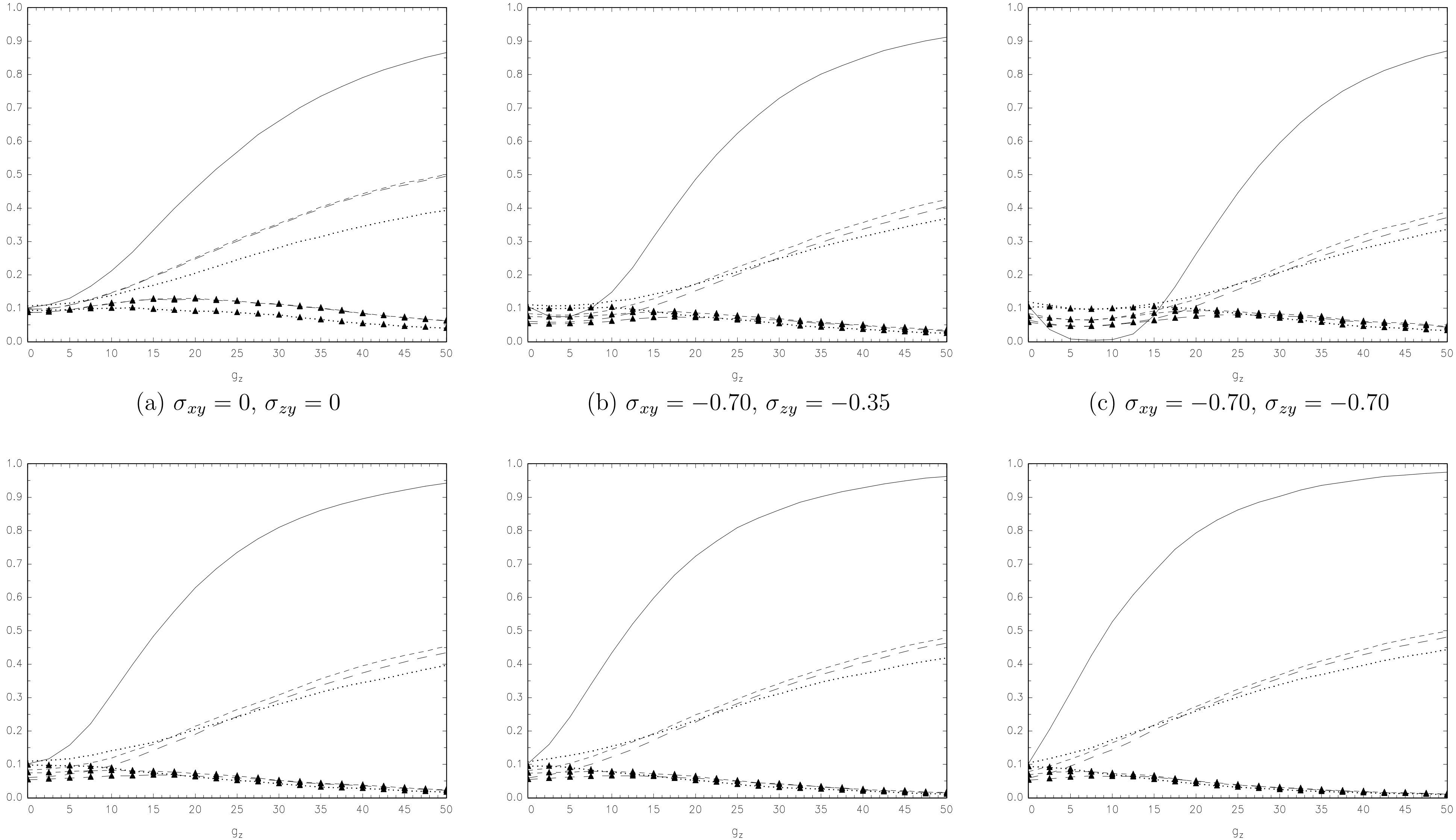


(e) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0.35$



(f) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$

Figure 6. Finite sample rejection frequencies of S_B (power) and t_u, Q, IV_{comb}, t_u^{pre}, Q^{pre}, IV_{comb}^{pre} (size): $T = 200$, $g_x = 0$, $c_x = c_z = 5$;
S_B: —— , t_u: - - - , Q: – – , IV_{comb}: $\cdots$, t_u^{pre}: - ▲ - , Q^{pre}: – ▲ – , IV_{comb}^{pre}: $\cdots$ ▲ $\cdots$



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

(b) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.35$

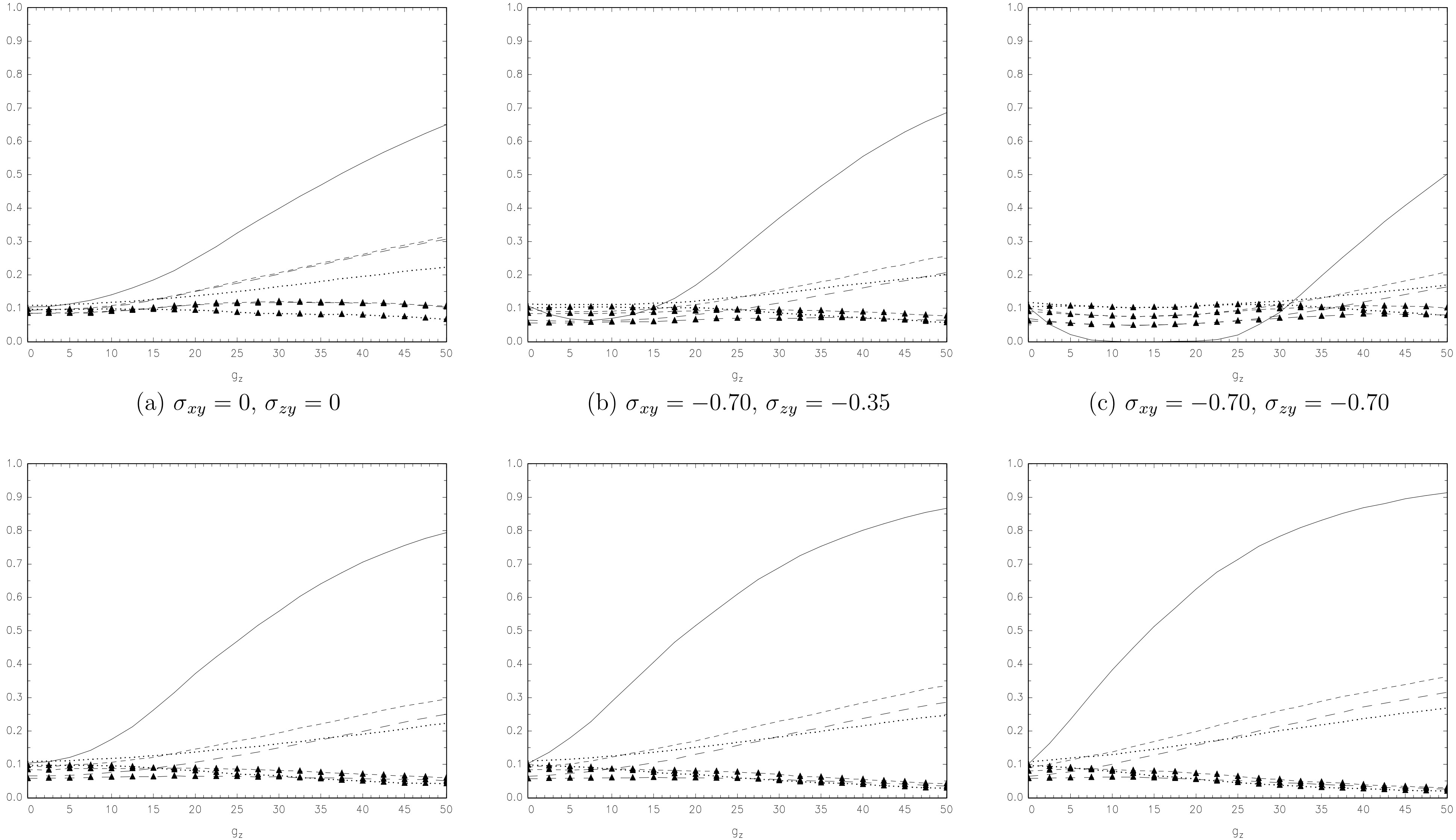
(c) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$

(d) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0$

(e) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0.35$

(f) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$

Figure 7. Finite sample rejection frequencies of S_B (power) and t_u, Q, IV_{comb}, t_u^{pre}, Q^{pre}, IV_{comb}^{pre} (size): $T = 200$, $g_x = 0$, $c_x = c_z = 10$; S_B: —— , t_u: - - - , Q: – – , IV_{comb}: $\cdots$, t_u^{pre}: - ▲ - , Q^{pre}: – ▲ – , IV_{comb}^{pre}: $\cdots$▲$\cdots$



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

(b) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.35$

(c) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$

(d) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0$

(e) $\sigma_{xy} = -0.70$, $\sigma_{zy} = 0.35$

(f) $\sigma_{xy} = -0.70$, $\sigma_{zy} = -0.70$

Figure 8. Finite sample rejection frequencies of S_B (power) and t_u, Q, IV_{comb}, t_u^{pre}, Q^{pre}, IV_{comb}^{pre} (size): $T = 200$, $g_x = 0$, $c_x = c_z = 20$;
S_B: ——, t_u: - - - , Q: – – , IV_{comb}: $\cdots$, t_u^{pre}: - ▲ - , Q^{pre}: – ▲ – , IV_{comb}^{pre}: $\cdots$ ▲ $\cdots$