Essex Finance Centre
Working Paper Series

Working Paper No29: 01-2018

Testing for Parameter Instability in Predictive Regression Models

“Iliyan Georgiev, David I. Harvey, Stephen J. Leybourne and A.M. Robert Taylor”

Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ
Web site: http://www.essex.ac.uk/ebs/

Testing for Parameter Instability in Predictive Regression Models*

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

[a]Department of Statistical Sciences, Università di Bologna

[b] School of Economics, University of Nottingham

[c]Essex Business School, University of Essex

Abstract

We consider tests for structural change, based on the *SupF* and Cramer-von-Mises type statistics of Andrews (1993) and Nyblom (1989), respectively, in the slope and/or intercept parameters of a predictive regression model where the predictors display strong persistence. The *SupF* type tests are motivated by alternatives where the parameters display a small number of breaks at deterministic points in the sample, while the Cramer-von-Mises alternative is one where the coefficients are random and slowly evolve through time. In order to allow for an unknown degree of persistence in the predictors, and for both conditional and unconditional heteroskedasticity in the data, we implement the tests using a fixed regressor wild bootstrap procedure. The asymptotic validity of the bootstrap tests is established by showing that the asymptotic distributions of the bootstrap parameter constancy statistics, conditional on the data, coincide with those of the asymptotic null distributions of the corresponding statistics computed on the original data, conditional on the predictors. Monte Carlo simulations suggest that the bootstrap parameter stability tests work well in finite samples, with the tests based on the Cramer-von-Mises type principle seemingly the most useful in practice. An empirical application to U.S. stock returns data demonstrates the practical usefulness of these methods.

Keywords: Predictive regression; persistence; parameter stability tests; fixed regressor wild bootstrap; conditional distribution.

JEL Classification: C12, C32, C58.

*We are very grateful to the Co-Editor, Oliver Linton, an Associate Editor and three anonymous referees for their helpful and constructive comments on earlier versions of this paper. Taylor gratefully acknowledges financial support provided by the Economic and Social Research Council of the United Kingdom under research grant ES/R00496X/1. 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 regression (hereafter PR) is a widely used tool in applied finance and economics. A leading example concerns whether future stock returns can be predicted by current information. In this context PR methods have been extensively utilised in studies of mutual fund performance, tests of the conditional CAPM and studies of optimal asset allocation; see Paye and Timmermann (2006,pp.274-275) and references therein. Predictors commonly considered for returns include the dividend yield, the term structure of interest rates, and default premia. It is often found that the posited predictor (e.g. the dividend yield) exhibits strongly persistent behaviour akin to that of a (near-) unit root autoregressive process, whilst the variable being predicted (e.g. the stock return) resembles a (near-) martingale difference sequence [m.d.s.].

Predictability tests which are asymptotically valid when the putative predictor is strongly persistent and driven by innovations which are correlated with the series being predicted (the latter is often thought to be the case; e.g., the stock price is a component of both the return and the dividend yield) have been proposed in Cavanagh *et al.* (1995), Campbell and Yogo (2006), Kostakis *et al.* (2015), Breitung and Demetrescu (2015), Elliott *et al.* (2015) and Jansson and Moreira (2006), *inter alia*. These approaches are all based on the maintained assumption that the coefficients of the PR model are constant over time. There is, however, a growing body of empirical evidence casting doubt on this assumption. Henkel *et al.* (2011), for example, find that return predictability in the stock market appears to be closely linked to economic recessions with dividend yield and term structure variables displaying predictive power only during recessions. Johannes *et al.* (2014) find strong empirical evidence of time-variation in the parameters of PRs for returns, including evidence of non-constant volatility. Timmermann (2008) argues that for most time periods stock returns are not predictable but that there are 'pockets in time' where evidence of local predictability is seen. Paye and Timmermann (2006) also cite a number of applied studies which find significant evidence of in-sample (ex post) predictability in returns data but yet find very weak evidence of out-of-sample (ex ante) predictability, and argue that a possible explanation is structural instability in the predictive relations involved.

Paye and Timmermann (2006) use a variety of well-known structural change tests designed against abrupt (deterministic) changes in a model's parameters to investigate the structural stability of PRs for stock returns related to structural breaks in the coefficients of state variables (including the lagged dividend yield, short interest rate, term spread and default premium) for a data-set of monthly stock returns for ten OECD countries. They find evidence of instability for the vast majority of these countries, arguing that the "Empirical evidence of predictability is not uniform over time and is concentrated in certain periods." *op.cit.* p.312. They also present simulation evidence into the size and power of the structural change tests they consider for the case where the predictors involved are $I(0)$ and a one-time break is allowed in the coefficient on a single predictor, and conclude in favour of the approach of Bai and Perron (1998,2003). A significant drawback of applying the Bai and Perron approach to the PR model, however, is that it is not asymptotically valid in cases where the predictive variables are (near-) unit root processes. Moreover, as argued by Cai *et al.* (2015,p.954) and the references therein, its focus on models of abrupt deterministic coefficient change might be considered unattractive in practice relative to tests designed for the case where the parameters of the PR are random and evolve smoothly over time. Indeed, using Bayesian model selection and averaging methods, Dangl and Halling (2012) conclude that time-variation in the coefficients of return prediction models is very important with a random walk coefficients model performing best in practice, quickly adapting to changes in environment. They also find evidence suggestive that predictability is linked to the business cycle. The Bai and Perron approach also requires that the variables in the PR do not display unconditional heteroskedasticity which would again appear to considerably limit their applicability for financial data; see, e.g., Johannes *et al.* (2014).

Our aim here is to address these shortcomings and develop structural change tests that can be more reasonably applied to empirically testing the constancy of the intercept and slope parameters in a PR model driven by heteroskedastic innovations. In earlier work, Georgiev *et al.* (2018) [GHLT hereafter], we investigated a variant of the stationarity test of Kwiatkowski *et al.* (1992) [KPSS] in the context of the PR model. This is a test of the instability of the regression intercept and can be viewed as a test against the alternative that the error in the PR model follows a near-unit root process. As such, GHLT interpret this as a test for spurious predictability. The present paper extends the work in GHLT to cover tests on all or a subset of the parameters of the PR model, not just the intercept, thereby allowing us to also investigate the constancy or otherwise of the slope parameter on the predictive regressor.

In the light of the arguments above, we consider parameter constancy tests based on the *SupF* type statistics of Andrews (1993) and the Cramer-von-Mises type statistics of Nyblom (1989). The former are designed for abrupt deterministic change models and the latter for (near-) unit root coefficient models. Although originally developed for *asymptotically stationary* regressors, Hansen (1992a) examines the large sample properties of these statistics for the case of pure unit root regressors, showing how these limits differ from the asymptotically stationary case. However, in the context of the PR model we need to go further and allow for the case where the predictive regressor is a near-unit root processes. Doing so introduces the considerable complication relative to the case of a pure unit root regressor that the limiting null distributions of the parameter constancy statistics depend on the local-to-unity (persistence) parameter of the putative predictor. In principle, this makes it very difficult to control the size of the tests given that this parameter is unknown in practice and cannot be consistently estimated.[1]

To resolve this problem we use bootstrap implementations of the parameter constancy tests which treat the putative predictor as a fixed regressor; i.e., the observed data on the predictor is used in calculating the bootstrap analogues of the structural change statistics. Because, as noted above, many economic and financial time series are thought to display non-stationary volatility and/or conditional heteroskedasticity, it is also important for our proposed bootstrap tests to

[1] Cai *et al.* (2015) also develop a test against smooth parameter variation in the parameters of the PR model based on a non-parametric L_2-type statistic. However, their proposed statistic requires the variables in the predictive regressor to be homoskedastic. We therefore do not consider their approach further here.

be (asymptotically) robust to these effects. To achieve this we use a heteroskedasticity-robust variant of the *fixed regressor bootstrap* approach proposed in Hansen (2000). We show that this approach yields asymptotically size-controlled tests, without requiring knowledge of the local-to-unity parameter or the form of any heteroskedasticity present, and delivers tests which are powerful against both forms of coefficient variation considered. Moreover, the bootstrap tests are also valid when the predictive regressors are asymptotically stationary or contain a mix of both asymptotically stationary and strongly persistent regressors. They are also valid for regressors whose marginal distributions are subject to structural change, meaning that rejection by the bootstrap tests can be unambiguously interpreted as evidence for structural instability in the slope coefficients of the PR, even where the predictors themselves display structural change.

Closely related to this paper, Hansen (2000) also applies the fixed regressor bootstrap to the Andrews (1993) and Nyblom (1989) statistics we consider here, and Paye and Timmermann (2006) include the fixed regressor bootstrap implementation of the Andrews (1993) test in their simulation study. Although Hansen (2000) employs assumptions that allow for pure and near unit root behaviour in the regressor variables, we demonstrate that his formal analysis needs an amendment. Hansen (2000) justifies bootstrap validity by claiming equivalence of the limiting distribution of the bootstrap parameter constancy statistics given the data and the *unconditional* limiting distribution of the original test statistics under the null. We show that this equivalence does not occur; in particular, the limits of the bootstrap statistics in his Theorems 5 and 6 on page 107 are both imprecisely stated when the predictive regressors are (near-) unit root processes. We establish that the fixed regressor bootstrap is nevertheless valid, at least in the PR set-up we consider, in the sense that the bootstrap parameter instability tests are asymptotically size-controlled. This is done by demonstrating that the limiting distributions of the bootstrap statistics, conditional on the data, are the same as the limiting null distributions of the corresponding statistics computed on the original data, *conditional* on the predictors.

The paper is organised as follows. Section 2 outlines our basic time-varying parameter PR model. To aid lucidity, we expound our approach through a single predictor variable whose innovations are serially uncorrelated. Generalisations to allow for multiple predictors and weak dependence are discussed in section 6. Section 3 outlines the structural change statistics and derives their asymptotic distributions. Section 4 details the fixed regressor wild bootstrap tests based on these statistics and establishes their asymptotic validity. The asymptotic local power of the bootstrap tests is examined in section 5, while section 7 presents Monte Carlo simulation results investigating their finite sample performance. An empirical application to monthly U.S. stock returns data is presented in Section 8. Section 9 concludes. Proofs appear in an appendix. Additional material relating to the limiting distributions of the statistics given in section 3 is provided in an accompanying on-line supplementary appendix.

2 The Predictive Regression Model with Structural Change

The basic PR model allowing for structural change that we consider for observed y_t is given by

$$y_t = \alpha_t + \beta_t x_{t-1} + \epsilon_{yt}, \qquad t = 1, ..., T \tag{1}$$

where x_t is an observed process, specified according to the data generating process [DGP]

$$x_t = \mu + s_{xt}, \qquad t = 0, ..., T \tag{2}$$

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

where $\rho_x := 1 - c_x T^{-1}$ with $c_x \geq 0$ such that x_t is a strongly persistent unit root or local-to-unit root autoregressive process with mean zero innovation process ϵ_{xt}. We let s_{x0} be an $O_p(1)$ variate. Exact conditions on the innovations ϵ_{yt} and ϵ_{xt} will be given in Assumption 1 below.

The DGP in (1) generalises the constant parameter PR model by allowing the intercept and slope coefficients to vary over time. To nest the constant parameter PR within (1) we formulate the time-varying intercept and slope coefficients as: $\alpha_t := (\alpha + as_{\alpha t})$ and $\beta_t := (\beta + bs_{\beta t})$. The parameter instability tests we discuss in this paper are, by construction, invariant to the values of α and β. However, in the context of a time-invariant PR (i.e., $\alpha_t = \alpha$, $\beta_t = \beta$) with near-unit root predictors it is usual to follow Cavanagh *et al.* (1995) and parameterise β to be local-to-zero at an appropriate rate; precisely, $\beta = gT^{-1}$ where g is some constant. This entails that under parameter constancy, and when g is non-zero, y_t is a near-m.d.s. process. Where β is fixed, as in Shin (1994), (1) should rightly be interpreted as a co-integrating regression because y_t will be a (near-) unit root process. However, because no particular parameterisation is needed for the theoretical results which follow (only for the PR interpretation of (1)) we do not directly impose a localisation on β. This is because in the case where x_t is (asymptotically) stationary no such standardisation of β is needed for a PR interpretation of (1).

In the context of (1) our focus will centre on testing the null hypotheses that the intercept and slope parameters are constant over time against the alternative that they vary over time through the sequences of associated time-varying coefficients, $s_{\alpha t}$ and $s_{\beta t}$. This can be done by testing the restrictions that $a = 0$ and $b = 0$ in (1). We will consider two possible mechanisms.

S: Stochastic Coefficient Variation

The first mechanism we consider for time variation in α_t and β_t in (1), in the spirit of Nyblom (1989), is one where $s_{\alpha t}$ and $s_{\beta t}$ follow (near-) unit root processes. That is,

$$\begin{bmatrix} s_{\alpha t} \\ s_{\beta t} \end{bmatrix} = \begin{bmatrix} \rho_\alpha & 0 \\ 0 & \rho_\beta \end{bmatrix} \begin{bmatrix} s_{\alpha t-1} \\ s_{\beta t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{\alpha t} \\ \epsilon_{\beta t} \end{bmatrix} \tag{4}$$

where $\rho_\alpha := 1 - c_\alpha T^{-1}$, $\rho_\beta := 1 - c_\beta T^{-1}$ with $c_\alpha \geq 0$, $c_\beta \geq 0$ which are unit root or local to unit root autoregressive processes.[2] Precise m.d.s.-based assumptions on the innovations $\epsilon_{\alpha t}$ and $\epsilon_{\beta t}$ will be given in Assumption 1. The coefficient processes are initialised at $s_{\alpha 0} = s_{\beta 0} = 0$.

[2] For (1) to be interpreted as a PR the parameter a should, paralleling the discussion surrounding β above, be localised as $a = g_\alpha T^{-1}$ under (4), otherwise y_t will be a (near-) unit root process.

In the context of (4), the PR in (1) reduces to a fixed coefficient model when $a = 0$ and $b = 0$. The intercept alone is random if $a \neq 0$ while $b = 0$. In this situation, if (1) is treated as a fixed coefficient regression model, it is then under-specified by an unobserved (local to) unit root autoregressive process; this is akin to the omission of a valid (local to) unit root predictive regressor, as studied in GHLT. If $a = 0$ while $b \neq 0$, treating (1) as a fixed coefficient regression model ignores the fact that the relationship between y_t and the predictive regressor x_{t-1} is not stable but is evolving through time. If $a \neq 0$ and $b \neq 0$, then both forms of mis-specification are present together when (1) is assumed to be a fixed coefficient model. In terms of hypothesis testing, then, we summarise these possibilities via the following taxonomy covering the null, H_0, and various alternatives, H^S, in the context of (1) and (4):

H_0 :	$a = 0,\ b = 0$	both intercept and x_{t-1} slope coefficient are fixed
H_1^S :	$a \neq 0,\ b = 0$	intercept only varies
H_x^S :	$a = 0,\ b \neq 0$	x_{t-1} slope coefficient only varies
H_{1x}^S :	$a \neq 0 \cup b \neq 0$	either the intercept or x_{t-1} slope coefficient, or both, vary.

N: Non-stochastic Coefficient Variation

The second mechanism we consider for time variation in α_t and β_t in (1) follows, among others, Andrews (1993) and is one where they are subject to abrupt changes which occur at a fixed number of deterministic points in the sample. For simplicity we will expound our analysis through the case of a one-time break, although the extension to allow for multiple such breaks is straightforward. However, where it is thought that multiple breaks are possible, the stochastic coefficient variation case might be considered a more natural framework; see also Remark 3 below. In the one-time break case $s_{\alpha t}$ and $s_{\beta t}$ are modelled as

$$s_{\alpha t} = s_{\beta t} = D_t(\lfloor \tau_0 T \rfloor) \tag{5}$$

where $D_t(\lfloor \tau T \rfloor) := \mathbb{I}(t > \lfloor \tau T \rfloor)$ with $\lfloor \tau T \rfloor$ denoting a generic shift point with associated break fraction τ, $\lfloor \cdot \rfloor$ the integer part of its argument and $\mathbb{I}(.)$ the indicator function. We take the true shift fraction τ_0 as unknown to the practitioner but to satisfy $\tau_0 \in \Lambda$, where $\Lambda = [\tau_L, \tau_U]$ with $0 < \tau_L < \tau_U < 1$. Here then, at time $\lfloor \tau_0 T \rfloor$, the intercept changes value from α to $\alpha + a$; the coefficient on x_{t-1} changes value from β to $\beta + b$. The corresponding taxonomy covering the null, H_0, and various alternatives, H^N, in the context of (1) and (5) is then:

H_0 :	$a = 0,\ b = 0$	both intercept and x_{t-1} slope coefficient are fixed
H_1^N :	$a \neq 0,\ b = 0$	intercept only shifts
H_x^N :	$a = 0,\ b \neq 0$	x_{t-1} slope coefficient only shifts
H_{1x}^N :	$a \neq 0 \cup b \neq 0$	either the intercept or x_{t-1} slope coefficient, or both, shift.

We conclude this section by detailing in Assumption 1 the conditions that we will place on the innovation vector $\epsilon_t := [\epsilon_{xt}, \epsilon_{yt}, \epsilon_{\alpha t}, \epsilon_{\beta t}]'$ in what follows, noting that only the assumptions pertaining to the leading two elements of ϵ_t are germane under scheme **N**. Some remarks follow.

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

(a) H and D_t are the 4×4 non-stochastic matrices

$$H := \begin{bmatrix} 1 & 0 & 0 & 0 \\ h_{21} & 1 & 0 & 0 \\ h_{31} & h_{32} & 1 & 0 \\ h_{41} & h_{42} & h_{43} & 1 \end{bmatrix}, \quad D_t := \begin{bmatrix} d_{1t} & 0 & 0 & 0 \\ 0 & d_{2t} & 0 & 0 \\ 0 & 0 & d_{3t} & 0 \\ 0 & 0 & 0 & d_{4t} \end{bmatrix}$$

with $h_{ij} \in \mathbb{R}$, $(i = 2,3,4,\ j = 1,2,3)$, such that HH' is strictly positive definite. The volatility terms d_{it} satisfy $d_{it} = d_i(t/T)$, where $d_i(\cdot) \in \mathcal{D}$, $\mathcal{D}^k := D_k[0,1]$ denoting the space of right continuous with left limit (càdlàg) $\mathbb{R}^k$-valued functions on $[0,1]$ equipped with the Skorokhod topology, are non-stochastic, strictly positive functions.

(b) e_t is a 4×1 vector m.d.s. with respect to a filtration $\mathcal{F}_t$, to which it is adapted, with conditional covariance matrix $\sigma_t := E(e_t e_t' | \mathcal{F}_{t-1})$ satisfying: (i) $T^{-1} \sum_{t=1}^{T} \sigma_t \xrightarrow{p} E(e_t e_t') = I_4$, $\xrightarrow{p}$ denoting convergence in probability as $T \to \infty$, and (ii) $\sup_t E\|e_t\|^{4+\delta} < \infty$ for some $\delta > 0$, where for any vector, x, $\|x\|$ denotes the usual Euclidean norm, $\|x\| := (x'x)^{1/2}$.

Remark 1. Assumption 1 implies that ϵ_t is a vector m.d.s. relative to $\mathcal{F}_t$, with conditional variance matrix $\Omega_{t|t-1} := E(\epsilon_t \epsilon_t' | \mathcal{F}_{t-1}) = (HD_t)\sigma_t(HD_t)'$, and time-varying unconditional variance matrix $\Omega_t := E(\epsilon_t \epsilon_t') = (HD_t)(HD_t)' > 0$.[3] Stationary conditional heteroskedasticity and non-stationary unconditional volatility are obtained as special cases with $d_i(\cdot) = d_i$, $i = 1,2,3,4$ (constant unconditional variance, hence only conditional heteroskedasticity), and $\sigma_t = I_4$ (so $\Omega_{t|t-1} = \Omega_t = \Omega(t/T)$, only unconditional non-stationary volatility), respectively. 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 a wide class of models for the behaviour of the variance matrix of ϵ_t (subject to the structure imposed by H), including single or multiple (co-) variance shifts, variances which follow a broken trend, and smooth transition variance shifts. Assumption 1(b) coincides with the m.d.s. conditions in Assumption 1 of Breitung and Demetrescu (2015), except that the cross product moment summability condition given there is not required as we do not allow ϵ_{xt} to be serially correlated at this stage. We will discuss extensions to allow for this in section 6.1 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 2. Assumption 1 permits correlation between the elements of ϵ_t through the elements h_{ij}, $i = 2,3,4$, $j = 1,2,3$, of the matrix H. In particular, where $h_{21} \neq 0$, then y_t and the innovations driving x_t, ϵ_{xt}, are correlated. □

Remark 3. Where $c_\alpha = c_\beta = 0$ such that $\rho_\alpha = \rho_\beta = 1$, Assumption 1 entails that $[s_{\alpha t},\ s_{\beta t}]'$ is a martingale of the form considered in Equation (2.1) of Nyblom (1989,p.224). This permits α_t and β_t to undergo either a deterministic or a random number of jumps of random magnitude,

[3] Notice that the assumptions that $E(e_t e_t') = I_4$ made in part (b)i and that the leading diagonal elements of H are unity involve no loss of generality.

with the number of jumps remaining (on average) a non-vanishing fraction of the sub-sample size, in every sub-sample. Where the (expected) number of jumps is lower than the sample size, they occur at random points in the sample. Moreover, as we allow $\epsilon_{\alpha t}$ and $\epsilon_{\beta t}$ to display non-constant unconditional variances these jumps can be such that they are much more likely to occur in certain parts of the sample than others allowing for clustering of the jumps. Where $c_\alpha > 0$, $c_\beta > 0$, $[s_{\alpha t},\ s_{\beta t}]'$ is a near-martingale and the coefficient processes α_t and β_t display long run mean reversion (towards α and β respectively). □

3 Parameter Constancy Tests

We first outline the structural change statistics that we will consider for testing parameter constancy in the PR in (1). We will then establish the large sample properties of these statistics.

3.1 Structural Change Test Statistics

S: Stochastic Coefficient Variation

To test H_0 against H^S we adopt the LM statistic of Nyblom (1989). Under certain conditions, including homoskedasticity and the requirement that $\rho_\alpha = \rho_\beta = 1$ in (4), then, conditional on x_t, this test statistic has a Locally Best Invariant (LBI) property. For testing H_0 against H^S_{1x}, the relevant LM statistic is given by

$$LM_{1x} := \frac{1}{T\hat{\sigma}^2}\sum_{i=1}^{T}\left(\sum_{t=1}^{i}\mathbf{a}_t'\hat{e}_t\right)\left(\sum_{t=1}^{T}\mathbf{a}_t\mathbf{a}_t'\right)^{-1}\left(\sum_{t=1}^{i}\mathbf{a}_t\hat{e}_t\right) \tag{6}$$

where $\mathbf{a}_t := [1\ x_{t-1}]'$, with $\hat{\sigma}^2 := T^{-1}\sum_{t=1}^{T}\hat{e}_t^2$, with $\hat{e}_t$ the OLS residual from the fitted regression

$$y_t = \hat{\alpha} + \hat{\beta}x_{t-1} + \hat{\beta}_0\Delta x_t + \hat{e}_t, \qquad t = 1, ..., T. \tag{7}$$

As in Shin (1994), (7) contains the additional regressor Δx_t, to account for the possibility of a non-zero correlation between ϵ_{xt} and ϵ_{yt} (which occurs when $h_{21} \neq 0$ in Assumption 1). The same will be needed in the context of the *SupF* statistics considered below.

We can also consider the corresponding single parameter LM statistics. These are given by

$$LM_1 := \frac{1}{T^2\hat{\sigma}^2}\sum_{i=1}^{T}\left(\sum_{t=1}^{i}\hat{e}_t\right)^2 \quad \text{and} \quad LM_x := \frac{1}{(T\hat{\sigma}^2\sum_{t=1}^{T}x_{t-1}^2)}\sum_{i=1}^{T}\left(\sum_{t=1}^{i}x_{t-1}\hat{e}_t\right)^2 \tag{8}$$

for the test statistics relating to the intercept alone and to the slope coefficient alone, respectively. Therefore LM_1 is appropriate for testing H^S_1, while LM_x is appropriate for testing H^S_x. The LM_1 statistic coincides with the statistic proposed in GHLT.

N: Nonstochastic Coefficient Variation

To test H_0 against H^N we use the *SupF* statistic of Andrews (1993). In a rather general, but asymptotically stationary setting, Andrews (1993) shows that a test based on this statistic has certain weak asymptotic local optimality properties against this form of parameter variation. For testing H_0 against H_{1x}^N in (1) and (5), this statistic is given by

$$SupF_{1x} := \sup_{\tau\in\Lambda} F(\tau), \quad F(\tau) := T\frac{\hat{\sigma}^2 - \hat{\sigma}^2(\tau)}{\hat{\sigma}^2(\tau)} \tag{9}$$

with $\hat{\sigma}^2$ defined as above, and $\hat{\sigma}^2(\tau) := T^{-1}\sum_{t=1}^T \hat{e}_t^2(\tau)$, $\hat{e}_t(\tau)$ from the fitted OLS regression

$$y_t = \hat{\alpha} + \hat{\alpha}^* D_t(\lfloor\tau T\rfloor) + \hat{\beta}x_{t-1} + \hat{\beta}^* D_t(\lfloor\tau T\rfloor)x_{t-1} + \hat{\beta}_0\Delta x_t + \hat{e}_t(\tau), \qquad t = 1, ..., T. \tag{10}$$

To test against H_1^N, exclude $D_t(\lfloor\tau T\rfloor)x_{t-1}$ from (10); denote the resulting statistic by $SupF_1$. For testing against H_x^N, $D_t(\lfloor\tau T\rfloor)$ is excluded from (10), and we denote this statistic by $SupF_x$.

Remark 4. The LM and *SupF* statistics are used to test the same null hypothesis, H_0, but differ in which alternative hypothesis they are directed towards. Still, as Hansen (1992a,p.325) points out they will "...tend to have power in similar directions..." The numerical results reported later in this paper accord with this view. Hansen argues that, as a result, the choice between the tests might be made on computational grounds and argues that this favours the LM statistic. He also argues that the purpose of the test is important and that if one is looking to test against a rapid change in regime then the *SupF* statistics would be appropriate, while "...if one is simply interested in testing whether or not the specified model is a good model that captures a stable relationship, the notion of martingale parameters is more appropriate, since it captures the notion of an unstable model that gradually shifts over time." *op. cit.* p.325. □

3.2 Asymptotic Distribution Theory

Under Assumption 1, the conditions of Lemma 1 of Boswijk *et al.* (2016) are satisfied such that

$$\left(T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\epsilon_t, T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\epsilon_s\epsilon_t'\right) \xrightarrow{w} \left(M_\eta(\cdot), \int_0^1 M_\eta(s)dM_\eta(s)'\right) \tag{11}$$

where $\xrightarrow{w}$ denotes weak convergence as $T\to\infty$, and $M_\eta(\cdot) := [M_{\eta x}(\cdot), M_{\eta y}(\cdot), M_{\eta\alpha}(\cdot), M_{\eta\beta}(\cdot)]'$ is a Gaussian martingale satisfying

$$\begin{bmatrix} M_{\eta x}(\cdot) \\ M_{\eta y}(\cdot) \\ M_{\eta\alpha}(\cdot) \\ M_{\eta\beta}(\cdot) \end{bmatrix} := H\begin{bmatrix} \int_0^\cdot d_1(s)dB_1(s) \\ \int_0^\cdot d_2(s)dB_2(s) \\ \int_0^\cdot d_3(s)dB_3(s) \\ \int_0^\cdot d_4(s)dB_4(s) \end{bmatrix}$$
$$= \begin{bmatrix} \{\int_0^1 d_1^2(s)\}^{1/2} & 0 & 0 & 0 \\ h_{21}\{\int_0^1 d_1^2(s)\}^{1/2} & \{\int_0^1 d_2^2(s)\}^{1/2} & 0 & 0 \\ h_{31}\{\int_0^1 d_1^2(s)\}^{1/2} & h_{32}\{\int_0^1 d_2^2(s)\}^{1/2} & \{\int_0^1 d_3^2(s)\}^{1/2} & 0 \\ h_{41}\{\int_0^1 d_1^2(s)\}^{1/2} & h_{42}\{\int_0^1 d_2^2(s)\}^{1/2} & h_{43}\{\int_0^1 d_3^2(s)\}^{1/2} & \{\int_0^1 d_4^2(s)\}^{1/2} \end{bmatrix} \times \begin{bmatrix} B_{\eta 1}(\cdot) \\ B_{\eta 2}(\cdot) \\ B_{\eta 3}(\cdot) \\ B_{\eta 4}(\cdot) \end{bmatrix}$$

with $B_{\eta i}(\cdot) := \{\int_0^1 d_i^2(s)\}^{-1/2}\int_0^\cdot d_i(s)dB_i(s)$, $i = 1,2,3,4$, and $[B_1(\cdot), B_2(\cdot), B_3(\cdot), B_4(\cdot)]'$ a standard Brownian motion in $\mathbb{R}^4$. We can also write $B_{\eta i}(\cdot) \stackrel{d}{=} B_i(\eta_i(\cdot))$, $i = 1,2,3,4$, where

$\eta_i(\cdot)$ denotes the variance profile $\eta_i(\cdot) := \{\int_0^1 d_i^2(s)\}^{-1} \int_0^{\cdot} d_i^2(s)ds$, $i = 1, 2, 3, 4$, such that $B_{\eta i}(\cdot)$ is a variance-transformed Brownian motion; see, for example, Davidson (1994, p.484). Under unconditional homoskedasticity, $\eta_i(s) = s$.

It will also prove convenient to define the Ornstein-Uhlenbeck [OU] type processes $B_{\eta 1, c_x}(r) := \int_0^r e^{-(r-s)c_x} dB_{\eta 1}(s)$, $M_{\eta\alpha, c_\alpha}(\cdot) := \int_0^r e^{-(r-s)c_\alpha} dM_{\eta\alpha}(s)$ and $M_{\eta\beta, c_\beta}(\cdot) := \int_0^r e^{-(r-s)c_\beta} dM_{\eta\beta}(s)$, for $r \in [0, 1]$, along with $M_{\eta x, c_x}(\cdot) := \{\int_0^1 d_1^2(s)ds\}^{1/2} B_{\eta 1, c_x}(\cdot)$ and its de-meaned analogue, $\bar{M}_{\eta x, c_x}(\cdot) := M_{\eta x, c_x}(\cdot) - \int_0^1 M_{\eta x, c_x}(s)ds$.

In order to examine the asymptotic local power properties of the tests we discuss we will specify H^S and H^N as local to H_0 by normalising the parameters a and b to be local-to-zero. The relevant normalisations are different for a and b and differ according to which form of coefficient variability is being considered. Specifically, under scheme **S** these are given by $a = g_\alpha T^{-1}$ in H_1^S and H_{1x}^S and $b = g_\beta T^{-3/2}$ in H_x^S and H_{1x}^S, while under scheme **N** these are given by $a = g_\alpha T^{-1/2}$ in H_1^N and H_{1x}^N and $b = g_\beta T^{-1}$ in H_x^N and H_{1x}^N. In each case g_α and g_β are fixed Pitman drift constants. Notice also that in these local settings H^S and H^N reduce to H_0 when $g_\alpha = g_\beta = 0$. In what follows, reference to these alternative hypotheses is understood to be made under these localisations, unless otherwise stated.

We now provide representations for the asymptotic distributions of the LM and *SupF* statistics under the local alternatives stated above. In Theorem 1 we do this for LM_{1x} and $SupF_{1x}$ in terms of matrix-valued processes. Alternative expressions for these limiting distributions in terms of scalar processes, together with those for the single parameter LM_1, LM_x, $SupF_1$ and $SupF_x$ statistics are provided in the on-line supplement.

Theorem 1. *Consider the model in (1), (2), (3) and let Assumption 1 hold. Then under the null hypothesis and the local alternatives outlined above,*

$$LM_{1x} \xrightarrow{w} \int_0^1 \mathbf{J}'(r)\{\mathbf{V}(1)\}^{-1}\mathbf{J}(r)dr$$

$$SupF_{1x} \xrightarrow{w} \sup_{r\in\Lambda} \left(\mathbf{J}(r)'\{\mathbf{V}(r) - \mathbf{V}(r)\mathbf{V}(1)^{-1}\mathbf{V}(r)\}^{-1}\mathbf{J}(r)\right),$$

where $\mathbf{J}(r) := \int_0^r \mathbf{A}(s)\, dY(s) - \mathbf{V}(r)\mathbf{V}(1)^{-1} \int_0^1 \mathbf{A}(s)\, dY(s)$ *and* $\mathbf{V}(r) := \int_0^r \mathbf{A}(s)\mathbf{A}'(s)\, ds$ *with* $\mathbf{A}(r) := [1, \bar{M}_{\eta x, c_x}(r)]'$ *and* $Y(r) := B_{\eta 2}(r) + \left\{\int_0^1 d_2^2(s)ds\right\}^{-1/2} \int_0^r Q(s)ds$. *Finally, the process* $Q(r)$, $r \in [0, 1]$ *is defined such that* $Q(r) := g_\alpha M_{\eta\alpha, c_\alpha}(r) + g_\beta M_{\eta\beta, c_\beta}(r) M_{\eta x, c_x}(r)$ *under scheme* **S**, *while under under scheme* **N**, $Q(r) := \{g_\alpha + g_\beta M_{\eta x, c_x}(r)\}\mathbb{I}(r \geq \tau_0)$.

Remark 5. The limit expressions given in Theorem 1 for the LM_{1x} and $SupF_{1x}$ statistics can be regarded as statistics of the *LM* and *SupF* type, respectively, in the context of a continuous-time least squares regression of $dY(r)$ on dr and $\bar{M}_{\eta x, c_x}(r)dr$. Under the null hypothesis of parameter stability, $Q(r) = 0$ for all r in these representations, while under the alternative hypotheses considered, the presence of parameter instability affects the limit distributions of both test statistics through the process $Q(\cdot)$ which is a function of the Pitman drifts, g_α and g_β. Although motivated under specific forms of instability, both statistics can therefore be seen to be sensitive to both of the considered alternatives. □

Remark 6. The representations in Theorem 1 for the limiting distributions of LM_{1x} and $SupF_{1x}$ depend, under both the null, H_0, and the local alternatives considered, on the local-to-unity parameter, c_x, characterising the degree of persistence in x_t. For LM_{1x} this dependence can be seen more clearly in the alternative representation of its limiting distribution in Corollary S.1 in the supplement. For $SupF_{1x}$, consider for simplicity the benchmark case of unconditional homoskedasticity in ϵ_t, and observe first that the limiting processes $\mathbf{J}(\cdot)$, $\mathbf{V}(\cdot)$ and $Q(\cdot)$ all depend on c_x. Under H_0, for fixed $r \in \Lambda$, dependence on c_x disappears from the distribution of $\{\mathbf{V}(r) - \mathbf{V}(r)\mathbf{V}(1)^{-1}\mathbf{V}(r)\}^{-1/2}\mathbf{J}(r) \stackrel{d}{=} N(0, I_2)$, both conditionally on $\{\mathbf{A}(s)\}_{s\in[0,1]}$ and unconditionally, because in this case $\mathbf{J}(\cdot)$ conditional on $\{\mathbf{A}(s)\}_{s\in[0,1]}$ is a zero-mean Gaussian process with covariance function $E\{\mathbf{J}(r_1)\mathbf{J}'(r_2)|\{\mathbf{A}(s)\}_{s\in[0,1]}\} = \mathbf{V}(r_1) - \mathbf{V}(r_1)\mathbf{V}(1)^{-1}\mathbf{V}(r_2)$ for $r_1 \leq r_2$. It follows that the limiting null distribution of $F(\tau)$ from (9) under unconditional homoskedasticity is $\chi^2(2)$ regardless of whether x_{t-1} is a pure unit root ($c_x = 0$), a near-unit root ($c_x > 0$), or even an asymptotically stationary process (see Andrews, 1993, for the latter). However, upon taking the supremum over $r \in \Lambda$, the distribution of the resulting functional $SupF_{1x}$ conditional on $\{\mathbf{A}(s)\}_{s\in[0,1]}$ depends on all the conditional covariances of $\{\mathbf{V}(r) - \mathbf{V}(r)\mathbf{V}(1)^{-1}\mathbf{V}(r)\}^{-1/2}\mathbf{J}(r)$ and not just on its trivial conditional variance, and so depends on c_x. This dependence carries over to the unconditional distribution of $SupF_{1x}$. □

Remark 7. It can also be seen from the representations given in Theorem 1 that the limiting distributions of the structural change statistics do not depend on any of the elements of the matrix H in Assumption 1, under either the constant parameter null hypothesis, H_0, or the local alternatives involving non-stochastic coefficient variation $\mathbf{N}$. They do, however, depend on any unconditional non-stationary volatility present in the innovations through the variance transformed Brownian motion $B_{\eta 2}(r)$ and the scaled variance transformed OU process $M_{\eta x, c_x}(r)$; that is, from any unconditional heteroskedasticity present in ϵ_{yt} and ϵ_{xt}. Where the local alternatives pertain to the stochastic coefficient variation scheme $\mathbf{S}$ in (4), these limiting distributions also depend on any unconditional heteroskedasticity present in $\epsilon_{\alpha t}$ when $g_\alpha \neq 0$ and in $\epsilon_{\beta t}$ when $g_\beta \neq 0$, and on the correlation between ϵ_{yt}, ϵ_{xt} and $\epsilon_{\alpha t}$, $\epsilon_{\beta t}$. □

Remark 8. The representations given in Theorem 1 fit with the generic representations given in Theorem 2 of Hansen (2000). Hansen (2000,p.98) gives a set of high-level conditions governing the weak convergence of the sample moments of the data and gives representations for the limiting distributions of the statistics under those conditions. The model set-up with associated assumptions that we consider here is, in the benchmark case of unconditional homoskedasticity in ϵ_t and of no correlation between ϵ_{yt} and ϵ_{xt} (i.e., $h_{21} = 0$), an example which satisfies Hansen's conditions and so we would expect our result to accord with his generic result. This is indeed seen to be so on noting that the processes $V(\cdot)$ and $\int_0^{\cdot} \mathbf{A}(s)\, dY(s)$ which appear in Theorem 1 coincide with the generic $M(\cdot)$ and $N(\cdot)$ processes, respectively, in Theorem 2 of Hansen (2000) under the specific conditions of the benchmark case outlined above. □

Remark 9. Where x_{t-1} is asymptotically stationary (in the sense of Definition 1 of Hansen, 2000), and the error term $d_{2t}e_{2t}$ is homoskedastic, the asymptotic null distributions of the LM_{1x}

and $SupF_{1x}$ statistics are, by Theorem 1 of Hansen (2000), of the form given in Equation (3.3) of Nyblom (1989,p.226) and Theorem 3 of Andrews (1993,p.838), respectively. More generally, consider the case where $T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}[1,x_{t-1}]'[1,x_{t-1}]$ and $T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}[1,x_{t-1}]'[1,x_{t-1}]d_{2t}^2e_{2t}^2$, $r\in[0,1]$, converge in probability to deterministic processes (say, $\widetilde{V}(\cdot)$ and $\{\int_0^1 d_2^2(s)\}^{1/2}\widetilde{V}_\eta(\cdot)$, respectively), which are continuous and, for $r>0$, positive definite. Further suppose that $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}[1,x_{t-1}]'d_{2t}e_{2t}$ converges weakly to a zero-mean Gaussian process (say, $\widetilde{J}_0(\cdot)$) with independent increments and variance function $\{\int_0^1 d_2^2(s)\}^{1/2}\widetilde{V}_\eta(\cdot)$, and let x_{t-1} be such that the inclusion of Δx_t in (7) eliminates the effects of $h_{21}d_{1t}e_{1t}$ from the residual, $\hat{e}_t$. Then the asymptotic null distributions of the LM_{1x} and $SupF_{1x}$ statistics are as given in Theorem 1, but with $\widetilde{J}_0$ and $\widetilde{V}$ replacing J and V, respectively.[4] The dependence of these distributions on (among other things) heteroskedasticity introduced by a general function $d_2(\cdot)$ of the form given in Assumption 1 would make the use of a bootstrap approximation desirable and, by the arguments of Theorems 5 and 6 of Hansen (2000), the fixed regressor wild bootstrap we outline in section 4 would be asymptotically valid. Because in this case the fixed regressor bootstrap statistics would converge to non-random distributions, the focus on conditioning which characterises the central results of this paper for the case of strongly persistent regressors becomes unnecessary. However, the key point is that the fixed regressor wild bootstrap implementations of the structural change tests we consider will be asymptotically valid regardless of whether x_t satisfies the conditions outlined in section 2 (the proof of which is given in section 4.1) or the generic conditions outlined above, and can therefore be validly used regardless of which of these two set-ups holds for x_t, or, when allowing for multiple predictors (see section 6.2), the case where both types are present; indeed, they are also asymptotically valid in cases where the degree of persistence of the variables in x_t changes over the sample. Moreover, as demonstrated in Theorems 5 and 6 of Hansen (2000), the fixed regressor bootstrap tests will also be asymptotically valid if the set of predictors includes regressors whose marginal distributions are subject to structural change of the form given in Section 4 of Hansen (2000). □

4 Fixed Regressor Wild Bootstrap Tests

As the results in the previous section show, implementing tests based on the LM and $SupF$ statistics will require us to address the fact that their limiting null distributions depend on any unconditional heteroskedasticity present in ϵ_{xt} and ϵ_{yt}, and on the persistence parameter c_x. To account for the former we employ a wild bootstrap procedure based on the residuals $\hat{e}_t$ of the fitted regression (7), while for the latter we use the observed outcome on $x:=[x_0,x_1,...,x_T]'$ as a fixed regressor when implementing the bootstrap procedure.

We now outline our fixed regressor wild bootstrap approach in Algorithm 1. To aid exposition we do so for the bootstrap tests based on the LM_{1x} statistic, but it should be entirely clear how the same approach can be applied to the LM_1, LM_x, $SupF_{1x}$, $SupF_1$ and $SupF_x$

[4] The corresponding limiting distributions under local alternatives can be obtained by appropriately modifying the limiting process $Q(\cdot)$ in Theorem 1.

statistics, with the resulting bootstrap analogues of these statistics correspondingly denoted by LM_1^*, LM_x^*, $SupF_{1x}^*$, $SupF_1^*$ and $SupF_x^*$, respectively.

Algorithm 1 (Fixed Regressor Wild Bootstrap):

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

(ii) Calculate the fixed regressor wild bootstrap analogue of LM_{1x} as outlined in section 3, but with y_t^* in place of y_t and with the regressor Δx_t omitted. Denote the resulting bootstrap statistic as LM_{1x}^*.

(iii) Define the corresponding p-value as $P_T^* := 1 - G_T^*(LM_{1x}^*)$, with $G_T^*(\cdot)$ denoting the conditional (on the original data) cumulative distribution function (cdf) of LM_{1x}^*. In practice, $G_T^*(\cdot)$ will be unknown, but can be simulated in the usual way.

(iv) The wild bootstrap test of H_0 at level ξ rejects if $P_T^* \leq \xi$.

Remark 10. Although $\hat{e}_t$ depends on g_α and/or g_β unless H_0 is true we will show in the next subsection that this does not translate into large sample dependence of LM_{1x}^* and $SupF_{1x}^*$ on these parameters. In the case of developing bootstrap tests based on the $SupF_{1x}$, $SupF_1$ and $SupF_x$ statistics and where it was thought that scheme **N** applied then one could also consider replacing $\hat{e}_t$ in step (i) of Algorithm 1 by the residuals $\hat{e}_t(\hat{\tau})$ where $\hat{\tau} := \arg\sup_{\tau\in\Lambda} F(\tau)$. This would not alter the large sample results which follow but in Monte Carlo experiments we found almost no difference between this approach and that outlined in Algorithm 1. □

Remark 11. Notice that in the bootstrap regression in step (ii) of Algorithm 1 we do not need to include Δx_t as an additional regressor. 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}. Also observe that we can assume that $\alpha = \beta = 0$ with no loss of generality when generating the bootstrap y_t^* data in step (i) because of the invariance of the residuals $\hat{e}_t$ to the values of α and β in (1). □

Remark 12. An alternative approach to accounting for unconditional heteroskedasticity in the context of the $SupF$ tests is to replace $F(\tau)$ in (9) with a corresponding robust Wald statistic based around a heteroskedastic-robust variance estimate; see White (1982). However, although the marginal limiting null distributions for these statistics, for a fixed value of τ, do not depend on any unconditional heteroskedasticity present in ϵ_{xt} and ϵ_{yt}, the suprema of the sequences of such statistics taken over all $\tau \in \Lambda$ do still depend, in general, on the heteroskedasticity, and hence a wild bootstrap would still be needed to obtain asymptotic size control. The limiting distributions of these sup-Wald statistics differ from those of the corresponding $SupF$ statistics under both the null and local alternatives and, as a result, their local power functions do not coincide. Similarly, one could also consider heteroskedasticity-corrected versions of the LM statistics, as discussed in Hansen (1992b), but the limiting distributions for these statistics are also not invariant to unconditional heteroskedasticity and so again a wild bootstrap would still be needed. In unreported finite sample simulations comparing these alternative approaches with those based on the tests outlined in section 3, we found neither approach to dominate the other overall in terms of size and power performance. □

4.1 Asymptotic Theory for the Bootstrap Tests

We first show that the limiting behaviour of the bootstrap statistics LM^*_{1x} and $SupF^*_{1x}$, conditional on the data, cannot be described in the standard terms of weak convergence in probability to a non-random distribution. Rather, to formulate a useful asymptotic result, a weaker convergence mode and a more general form of the limit are required. Using the concept of weak convergence of random measures, we demonstrate that the distributions of LM^*_{1x} and $SupF^*_{1x}$, given the data, converge to the *random* distributions which obtain by conditioning the limiting null distributions given in Theorem 1 on the weak limit B_1 of $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} e_{1t}$. Second, we establish that under H_0 and a strengthening of Assumption 1, the distributions of the LM_{1x} and $SupF_{1x}$ statistics, conditional on $x := [x_0, x_1, ..., x_T]'$, converge weakly to the same random distributions referred to above. This result allows us to establish the asymptotic validity of our bootstrap test. As in GHLT, in order to proceed 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) $\{e_{2:4,t}\}_{t=-\infty}^{\infty}$*, with* $e_{2:4,t} := [e_{2t}, e_{3t}, e_{4t}]'$*, 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* $\{e_{1s}\}_{s=-\infty}^{\infty}$ *and* $\{e_{2:4,s}\}_{s=-\infty}^{t}$*, respectively, and* $\mathcal{X} \vee \mathcal{F}_t$ *denotes the smallest* σ*-algebra containing both* $\mathcal{X}$ *and* $\mathcal{F}_t$*.*

(c) The initial value $s_{x,0}$ *is measurable w.r.t.* $\mathcal{X}$ *(in particular, it could be a fixed constant).*

Remark 13. A detailed discussion of the implications of Assumption 2 is given in GHLT to which we refer the reader. Assumption 2 enables us to invoke a conditional (on x) functional central limit theorem, together with a bootstrap analogue of that result for $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} y_t^*$ conditional on all of the data (x and $y := [y_1, ..., y_T]'$). Taken together with further results on conditional convergence to stochastic integrals adapted from GHLT, these results allow us to obtain the limiting distributions of the original statistics LM_{1x} and $SupF_{1x}$, conditional on x, together with the limiting distributions of the corresponding bootstrap LM^*_{1x} and $SupF^*_{1x}$ statistics from Algorithm 1, conditional on the data. These are now reported in Theorem 2 and underlie the validity of our bootstrap approach. □

Theorem 2. *Consider the model in (1), (2), (3) and let Assumption 2 hold. Under the null hypothesis and under the same local alternatives as were considered in the context of Theorem 1, the following converge jointly as* $T \to \infty$*, in the sense of weak convergence of random measures on* $\mathbb{R}$*:* $LM_{1x}|\, x \xrightarrow{w} \int_0^1 \mathbf{J}'(r)\{\mathbf{V}(1)\}^{-1}\mathbf{J}(r)dr\Big|\, B_1$*, and* $LM^*_{1x}|\, x, y \xrightarrow{w} \int_0^1 \mathbf{J}_0'(r)\{\mathbf{V}(1)\}^{-1}\mathbf{J}_0(r)dr\Big|\, B_1$*, where* $\mathbf{J}_0(r) := \int_0^r \mathbf{A}(s)dY(s)$*,* $r \in [0,1]$*, with* $\mathbf{J}$*,* $\mathbf{V}$*,* $\mathbf{A}$ *and* Y *defined in Theorem 1. Again in* *the sense of weak convergence of random measures on* $\mathbb{R}$*, the following converge jointly as* $T \to \infty$*:* $SupF_{1x}|\, x \overset{w}{\to} \sup_{r\in\Lambda} \left(\mathbf{J}'(r)\{\mathbf{V}(r) - \mathbf{V}(r)\mathbf{V}(1)^{-1}\mathbf{V}(r)\}^{-1}\mathbf{J}(r)\right)\Big|\, B_1$ *and* $SupF^*_{1x}|\, x, y \overset{w}{\to} \sup_{r\in\Lambda} \left(\mathbf{J}_0'(r)\{\mathbf{V}(r) - \mathbf{V}(r)\mathbf{V}(1)^{-1}\mathbf{V}(r)\}^{-1}\mathbf{J}_0(r)\right)\Big|\, B_1$*.*

Remark 14. For the precise meaning of joint weak convergence of random measures, we refer the reader to the Appendix and to the discussion on this point in section 4.3 of GHLT. The concept is weaker than weak convergence in probability, although it reduces to the latter when the limit distribution is non-random. Nevertheless, joint weak convergence of random measures implies convergence of the (conditional) distribution functions in a way that is still sufficient in order to yield consistency of the bootstrap in the usual p-value sense, as we will subsequently show in Corollary 1 below. □

Remark 15. Under the null hypothesis, the process $\mathbf{J}$ coincides with the process $\mathbf{J}_0$ whose form is invariant as to which of the null and local alternatives considered in this paper holds. As a result, the limiting distributions of the bootstrap statistics are the same under both the null and local alternatives and, moreover, coincide with the limiting null distributions, conditional on B_1, of the corresponding original test statistics. □

Remark 16. As discussed in Remark 6, in the case of unconditional homoskedasticity, the random variable $\mathbf{J}_0'(r)\{\mathbf{V}(r) - \mathbf{V}(r)\mathbf{V}(1)^{-1}\mathbf{V}(r)\}^{-1}\mathbf{J}_0(r)$ conditional on B_1 has a $\chi^2(2)$ distribution for every fixed $r \in \Lambda$ and, in particular, is independent of B_1. Nevertheless, even in this case, the conditional limiting null distribution of $SupF_{1x}$ and $SupF^*_{1x}$ is genuinely random (non-degenerate). This is so because the non-contemporaneous autocovariances of $\{\mathbf{V}(r) - \mathbf{V}(r)\mathbf{V}(1)^{-1}\mathbf{V}(r)\}^{-1/2}\mathbf{J}_0(r)$ conditional on B_1 depend on $\mathbf{V}$, and thus, on B_1 which is random. As a result, upon taking the supremum over $r \in \Lambda$, the distribution of the functional obtained, conditional on B_1, still depends on B_1 and is, therefore, random. Regarding LM_{1x} and LM^*_{1x}, the randomness of their conditional limiting null distributions is even more obvious because, even for fixed $r \in \Lambda$, the distribution of $\mathbf{J}_0'(r)\{\mathbf{V}(1)\}^{-1}\mathbf{J}_0(r)$ given B_1 is not independent of B_1, as $\mathbf{V}(1)$ is not the conditional variance of $\mathbf{J}_0(r)$. □

Remark 17. With a slight abuse of terminology, we could think of the random distributional limits of $SupF_{1x}$ and $SupF^*_{1x}$ (and likewise, of LM_{1x} and LM^*_{1x}) as random draws from a family of distributions indexed by B_1. Such random draws are distinct from the non-random mixture distribution obtained by averaging the family of distributions over B_1. Since the limit of $SupF^*_{1x}$ in Theorem 2 is distinct from this mixture distribution, it follows that the mixture distribution cannot be a weak limit in probability of $SupF^*_{1x}$, because weak convergence in probability implies convergence to the same limit also in the mode employed in Theorem 2. Furthermore, as the limits in Theorem 2 are invariant to the value of h_{21}, and our unconditionally homoskedastic case with $h_{21} = 0$ satisfies Assumption 2 of Hansen (2000) (see also Example 3 therein), we can conclude that the part of Theorem 6 in Hansen (2000) asserting the weak convergence in probability of Hansen's counterpart of $SupF^*_{1x}$ to the *un*conditional (and hence, non-random mixture) null limit distribution of $SupF_{1x}$ given in Theorem 1, is not correct. The same error appears in Theorem 3 of Cavaliere and Taylor (2006,p.626) who discuss fixed regressor wild bootstrap implementations of the Shin (1994) tests for the null of co-integration. Nevertheless, the ultimate claim in Hansen (2000), Corollary 2, that the bootstrap p-values under H_0 are asymptotically uniformly distributed (and, thus, that the fixed regressor wild bootstrap is asymptotically valid in this sense) can still be shown to hold true for the testing problem considered in this paper, though as a consequence of our Theorem 2 above (see Corollary 1 below). By similar considerations, the fixed regressor wild bootstrap implementations of the Shin (1994) tests in Cavaliere and Taylor (2006) could be shown to be asymptotically valid in the same sense. □

As we have seen in Theorem 2, the bootstrap statistics LM_{1x}^* and $SupF_{1x}^*$, conditional on the data, and the original statistics LM_{1x} and $SupF_{1x}$, conditional on x, share the same asymptotic distribution under the null hypothesis. We can obtain as an implication, now formalised in Corollary 1, that the bootstrap tests based on LM_{1x}^* and $SupF_{1x}^*$ are asymptotically valid. We state the result for LM_{1x} and $SupF_{1x}$ but the same conclusions hold for LM_1, LM_x, $SupF_1$ and $SupF_x$. As usual, validity is formulated in terms of bootstrap p-values.

Corollary 1. *Let the conditions of Theorem 2 hold. Then, under H_0, as $T \to \infty$, $P_{T,LM}^* := P^*(LM_{1x}^* \leq LM_{1x}) \xrightarrow{w} U[0,1]$ and $P_{T,F}^* := P^*(SupF_{1x}^* \leq SupF_{1x}) \xrightarrow{w} U[0,1]$.*

The practical implication of Corollary 1 is that comparison of one of the original statistics, for example LM_{1x}, with a ξ level empirical bootstrap critical value (calculated as the upper tail ξ percentile from the order statistic formed from B independent simulated bootstrap LM_{1x}^* statistics), which we will denote by $cv_{\xi,B}$, will result in a bootstrap test that under H_0 will have asymptotic size that for sufficiently large B will be as close as desired to the given nominal level ξ. Size in this context is understood to mean the rejection frequency in a thought experiment where the bootstrap test is applied to a large number of data samples constituting different realisations of the regressor $\{x_t\}$. This is distinct from the interpretation of the stronger results (also derived in the proof of Corollary 1) that $P_{T,LM}^*|x \xrightarrow{w}_p U[0,1]$ and $P_{T,F}^*|x \xrightarrow{w}_p U[0,1]$ under H_0, in the sense of weak convergence in probability; these results can be interpreted as also establishing the asymptotic validity of the bootstrap for fixed realisations of $\{x_t\}$. Under local alternatives $cv_{\xi,B}$ will remain as under H_0 (at least in the limit), while the distribution of LM_{1x} conditional on x will vary with g_α and g_β and so asymptotic local power of the bootstrap tests will be a function of those drift parameters. In what follows, as a matter of shorthand notation, we will denote by LM_{1x}^B the fixed regressor wild bootstrap procedure outlined in Algorithm 1, whereby the original statistic is compared to its empirical bootstrap critical value, $cv_{\xi,B}$.

5 Asymptotic Local Power

We now turn to a consideration of the asymptotic local power of the fixed regressor wild bootstrap procedures. In accordance with the interpretation given to the results in Corollary 1 above, we focus on asymptotic power understood as the rejection rate in a thought experiment with a large number of different realisations of the process B_1. We simulate the functionals in the limit distributions using 3000 Monte Carlo replications with different Brownian motion processes in each replication, approximated as random walks with $IID\, N(0,1)$ increments over a grid of 1000 points. For each replication, the simulated limit bootstrap critical value for $\xi = 0.10$ is obtained by simulating the appropriate bootstrap limit distribution using $B = 499$ bootstrap replications, conditioning on the simulated B_1 for that Monte Carlo replication.

In calculating asymptotic powers, in D_t we abstract from any role that non-stationary volatility plays by setting $d_{it} = 1$, for all i and t. We induce a correlation of -0.8 between ϵ_{xt} and ϵ_{yt} by setting $h_{21} = -4/3$; the other non-diagonal elements of H are set to 0. We also set $c_x = 10$. As regards the various alternatives, using a 30-step grid of values denoted g between 0 and 50, under stochastic parameter variation, **S**, we have in H^S: $g_\alpha = 3g/5$ for H_1^S, $g_\beta = 3g$ for H_x^S and $g_\alpha = 3g/5$, $g_\beta = 3g$ for H_{1x}^S and we consider $c_\alpha = c_\beta = \{0, 10\}$. Under non-stochastic parameter variation, **N**, we have in H^N: $g_\alpha = g/5$ for H_1^N, $g_\beta = g$ for H_x^N and $g_\alpha = g/5$, $g_\beta = g$ for H_{1x}^N and we consider the break fractions $\tau_0 = \{1/2, 3/4\}$ with $\tau_L = 0.1$ and $\tau_U = 0.9$. Here, the strength of the alternatives increases with g, the null being true for $g = 0$.

Figures 1 (a)-(c) report results for the stochastic parameter variation of H^S with $c_\alpha = c_\beta = 0$. In Figure 1 (a) the alternative is H_1^S (intercept variation only). Here it might be expected that LM_1^B would provide most power. However, there is very little to choose between this procedure and LM_{1x}^B, $SupF_1^B$ and $SupF_{1x}^B$. What is noticeable is that $SupF_x^B$ and especially LM_x^B, the two procedures that do not permit intercept variation of either type, perform significantly worse those those that do. Figure 1 (b) shows results for the alternative H_x^S (slope parameter variation) where LM_x^B might be expected to perform best. Here there is little difference between this procedure and LM_{1x}^B, $SupF_x^B$ and $SupF_{1x}^B$. We also observe that LM_1^B and $SupF_1^B$ perform much worst, with LM_1^B being least powerful of all. In Figure 1 (c) the alternative is H_{1x}^S (intercept and slope parameter variation). The two best procedures are LM_{1x}^B and $SupF_{1x}^B$ and there is little to choose between them. None of the other procedures performs particularly poorly, however. Figures 1 (d)-(f) repeat the same analysis with $c_\alpha = c_\beta = 10$. The powers of all procedures are now lower than when $c_\alpha = c_\beta = 0$, as would be expected. Otherwise, broadly speaking, the comments made for Figures 1 (a)-(c) apply here also.

In Figures 2 (a)-(c) we give results for the non-stochastic parameter variation of H^N for a mid-sample break, $\tau_0 = 1/2$. For Figure 2 (a) the alternative is H_1^N (intercept variation only). While we might expect $SupF_1^B$ to provide most power, it is clear that this role is actually fulfilled by LM_1^B, followed by $SupF_1^B$ and LM_{1x}^B, and then $SupF_{1x}^B$. As regards LM_x^B and $SupF_x^B$, the procedures that do not permit intercept variation, their power is again very low in comparison to the others. Figure 2 (b), where the alternative is H_x^N (slope parameter variation) reveals LM_x^B to be the best performing procedure, outperforming $SupF_x^B$. Here it is the power of LM_1^B and $SupF_1^B$ that are the lowest by some margin. In Figure 2 (c) the alternative is H_{1x}^N (intercept and slope parameter variation) and we see that the best procedure is LM_{1x}^B, followed by $SupF_{1x}^B$. The others have noticeably lower power compared to these two, though none of them performs badly. The analysis is repeated in Figures 2 (d)-(f) for a late break, $\tau_0 = 3/4$. Figure 2 (d), where the alternative is H_1^N (intercept variation), reveals that all the procedures that include this alternative now have fairly similar power; the power advantage previously evident for LM_1^B over $SupF_1^B$ is no longer in evidence, with both showing similar levels of power. Likewise, in Figure 2 (e) under the alternative H_x^N (slope parameter variation), we see that LM_x^B and $SupF_x^B$ also now have similar power levels. For the alternative H_{1x}^N (intercept and slope parameter variation) in Figure 2 (f), $SupF_{1x}^B$ generally appears more powerful than LM_{1x}^B, thereby reversing the previous ranking. Once more, we see that procedures which exclude parameter variation (of either type) perform badly when it is present in the alternative.

Summarising the findings of Figures 1 and 2, what is clear throughout is that procedures which incorrectly exclude the possibility of parameter variation associated with a particular regressor when it is present in the alternative in either form will lose power compared to those procedures that do permit one or other form of variation in that parameter. This is not really surprising. What is perhaps more surprising is that employing a procedure that specifies the correct form of parameter variation for a given alternative (i.e. stochastic or non-stochastic) does not always yield higher power than the corresponding procedure which specifies the incorrect form. In fact, the incorrectly specified procedure may have the higher power, as seen most obviously in the context of non-stochastic variation when the break fraction is $\tau_0 = 1/2$; here the LM-based procedures are consistently more powerful than their $SupF$-based counterparts.

6 Extensions

6.1 Weak Dependence

Thus far we have assumed that the noise, ϵ_{xt}, driving x_t is serially uncorrelated, by virtue of e_t being a m.d.s. More generally we might consider a linear process assumption for ϵ_{xt} of the form

$$\epsilon_{xt} = \sum_{i=0}^{\infty} \theta_i v_{x,t-i} \tag{12}$$

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 under this structure ϵ_{yt} remains uncorrelated with the lagged increments of x_t at all lags.

In this case, it may be shown that the limiting results given in this paper would continue to hold provided in (7) and (10) we add in the regressors $\Delta x_{t-1}, ..., \Delta x_{t-p}$ 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 above.[5] Similarly to Breitung and Demetrescu (2015), we would also need to restrict the amount of serial dependence allowed in the conditional variances via

[5] These regressors would not need to be added to the bootstrap analogues of (7) and (10) because the $\hat{e}_t$ used to construct y_t^* are free of any effects arising from weak dependence in ϵ_{xt}.

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. As is standard in the PR 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.

6.2 Multiple Predictors and Deterministic Components

The parameter constancy tests developed in the context of (1)-(3) with a single predictive regressor, x_{t-1}, and an intercept can be straightforwardly generalised to the case where the PR contains multiple predictors and/or a general deterministic component of the form considered in section 3.2 of Breitung and Demetrescu (2015).

Specifically, we may consider the case where the deterministic component in (1) is of the form $\alpha_t + \boldsymbol{\tau}'\mathbf{f}_t$, with α_t specified as before, and where $\mathbf{f}_t$ is as defined in section 3.2 of Breitung and Demetrescu (2015), but is such that it does not span the space of a constant; an obvious example is the linear trend case which obtains for $\mathbf{f}_t := t$. To allow for multiple predictors, replace x_{t-1} in (1) by the $k \times 1$ vector of predictive regressors as $\mathbf{x}_{t-1} := (x_{1,t-1}, ..., x_{k,t-1})'$ where each $x_{i,t}$ is generated by equations of the form given in (2) and (3), and where the former can also include the additional deterministic variables in $\mathbf{f}_t$. We would then correspondingly construct the *LM* structural instability statistics (which could be for single or joint parameter restrictions) with the residuals $\hat{e}_t$ now obtained from the regression of y_t onto and intercept, $\mathbf{f}_t$, $\mathbf{x}_{t-1}$ and $\Delta\mathbf{x}_{t-1}$ (and lags of $\Delta\mathbf{x}_{t-1}$ in the case considered in Section 6.1) and setting $\mathbf{a}_t := [1, \mathbf{x}_{t-1}']'$ in the calculation of (6). The bootstrap analogues of these statistics discussed in section 4 would use the residuals from the regression of y_t^* (the wild bootstrap analogue of y_t) onto an intercept, $\mathbf{f}_t$ and $\mathbf{x}_{t-1}$. For the *SupF*-type statistics the additional set of residuals $\hat{e}_t(\tau)$ needed to compute $F(\tau)$ in (9) are obtained from the regressions above but augmented with $D_t(\lfloor \tau T \rfloor)$ and/or $D_t(\lfloor \tau T \rfloor)\mathbf{x}_{t-1}$ and computed for each possible τ. For both the *LM* and *SupF*-type statistics, doing so alters the form of the limit distributions given in Theorem 1, but would not alter the primary conclusion given in Corollary 1, that the fixed regressor wild bootstrap implementation of the instability tests are asymptotically valid. In particular, the process $\mathbf{A}(\cdot)$ along with the de-meaned and tied-down Brownian-based processes which appear in Theorem 1 would need to be appropriately re-defined to the deterministic component being considered, $\mathbf{A}(\cdot)$ would now contain k OU derived processes, analogous to $\bar{M}_{\eta x, c_x}(\cdot)$, corresponding to each of the k elements of $\mathbf{x}_{t-1}$, while $Q(\cdot)$ would also now contain additional terms, analogous to $M_{\eta\beta, c_\beta}(\cdot)M_{\eta x, c_x}(\cdot)$ under scheme $\mathbf{S}$ and $M_{\eta x, c_x}(\cdot)$ under scheme $\mathbf{N}$, corresponding to each of the k elements of $\mathbf{x}_{t-1}$.

7 Finite Sample Size and Power

We now evaluate the finite sample size and power properties of the bootstrap procedures, on average over different realisations on x. We simulate the DGP (1)-(3) where we set $\mu = \alpha =$

$\beta = 0$, $s_{x0} = 0$ and generate $e_t \sim IID\, N(0, I_4)$ for a sample size of $T = 100$.[6] The simulations are again conducted using 3000 Monte Carlo replications, $B = 499$ bootstrap replications, and setting $\xi = 0.10$. No lagged Δx_t terms are incorporated into any fitted regression model for y_t.

In order to meaningfully compare finite sample results with the homoskedastic-case asymptotic results of the previous section, in the simulation DGPs we first employ exactly the same constellation of parameter settings as underpinned our reported asymptotic results. Figures 3 and 4 report our results for the finite sample analogues of Figures 1 and 2. Throughout, it is seen that each procedure has empirical size near to the nominal 0.10 level. It is also clear throughout that the finite sample powers generally bear a strong resemblance to their asymptotic counterparts in terms of the relative behaviour of the bootstrap procedures, and hence the comments given in the previous section apply here also (some discrepancies are simply due to small finite sample size differences). In absolute terms, the finite sample powers tend to be slightly lower than their asymptotic counterparts, although this is hardly noticeable in the cases of alternatives with non-stochastic parameter variation (Figure 4).

We next consider the impact of unconditional heteroskedasticity, investigating the finite sample size and power of our bootstrap procedures when two of the error processes, those for ϵ_{xt} and ϵ_{yt} are subject to a contemporaneous single break in volatility of equal magnitude. Specifically, we again simulate the DGP (1)-(3) with $T = 100$ letting $d_{it} = 1$ for $t \leq \lfloor \tau_{0h} T \rfloor$ and $d_{it} = \sigma$ for $t > \lfloor \tau_{0h} T \rfloor$, $i = 1, 2$, with $\tau_{0h} = \{1/2, 3/4\}$ and we consider $\sigma = \{4, 1/4\}$ thus allowing for both upward and downward volatility shifts, with the chosen magnitudes being substantial for illustrative purposes. The other simulation DGP settings are as in Figures 3 and 4, however for brevity we now only consider a subset of the values for g given by $g = \{0, 15, 35\}$. The results are shown in Tables 1 (a) and (b); these include the previously-considered homoskedastic case (obtained by setting $\sigma = 1$) as a benchmark for sizes and powers (note that in the tables, *SupF* is abbreviated to *SF*). Table 1 (a) considers the stochastic parameter variation of H^S, while Table 1 (b) considers the non-stochastic parameter variation of H^N. Size results are reported only in Panel A of Table 1 (a), to avoid unnecessary duplication.

In Panel A of Table 1 (a) the alternative is H_1^S (intercept variation). Beginning with empirical sizes of our procedures ($g = 0$), we see that heteroskedasticity has only a modest effect when compared to the benchmark homoskedastic case (particularly for the *LM*-based procedures). This suggests that the wild bootstrap is performing reasonably well in reproducing the patterns of heteroskedasticity present. Turning to finite sample power, in general terms we see that the upward volatility shift considered significantly decreases powers relative to the benchmark homoskedastic powers, while the downward shift considered has the opposite effect. These effects are observed for both volatility break timings considered. An examination of the power levels between the procedures reveals that under heteroskedasticity, the patterns of relative powers are generally similar to those observed in the homoskedastic case. For an alternative of H_x^S (slope parameter variation), from Panel B it appears that both upward and

[6] We also ran simulations for $T = 200$. These results were little different from those discussed here for $T = 100$ and so are omitted in the interests of brevity. These results can be obtained from the authors on request.

downward volatility shifts can lead to a decrease in power when compared to the homoskedastic benchmark (although this effect is rather small for the downward shift). The patterns of relative power levels between the procedures again largely mimic what we see under homoskedasticity. As might be expected, under the alternative H_{1x}^S (intercept and slope parameter variation), Panel C shows that the effect of a volatility shift on test power involves a mixture of the effects seen under H_1^S and H_x^S. In very general terms, the volatility effect on LM_1^B and $SupF_1^B$ under H_{1x}^S is most similar to that for H_1^S, the impact on LM_x^B and $SupF_x^B$ under H_{1x}^S is closer to that for H_x^S, while the effect of the volatility shift on LM_{1x}^B and $SupF_{1x}^B$ for H_{1x}^S is a hybrid of the effects seen under H_1^S and H_x^S. In Table 1 (b), where the alternatives are H_1^N, H_x^N and H_{1x}^N, volatility shifts are seen to have qualitatively similar effects on power, relative to the homoskedastic case, to those seen in Table 1 (a) for H_1^S, H_x^S and H_{1x}^S, respectively.

8 An Empirical Application

To illustrate how our proposed instability test procedures may be used in practice, we apply them to the U.S. annual equity series analysed in Welch and Goyal (2008), which is updated to cover the period 1926-2015 ($T = 90$) and is available at `http://www.hec.unil.ch/agoyal/`. Our y_t variables are R_t, the log of the total return (including dividends) on the S&P 500 stock market index from year $t-1$ to t, and EP_t, the equity premium, which subtracts the corresponding risk-free rate (the Treasury Bill rate) from R_t. The x_t predictor variables (in each case included in the bivariate PR with a one-period lag) are: the dividend yield, DY_t, defined as the difference between the log of dividends and the log of one-period lagged prices; the dividend payout ratio, DE_t, defined as the difference between the log of dividends and the log of earnings; and the long term rate of returns, LTR_t, the long term rate on government bonds. While 1926-2015 represents the full sample period, we also consider the sub-sample 1926-2007, which pre-dates the global financial crisis, allowing analysis of any potential differences when excluding the more recent years of instability.

The results are shown in Table 2(a). The main entries are the computed values of the LM_1^B, LM_x^B, LM_{1x}^B, $SupF_1^B$, $SupF_x^B$ and $SupF_{1x}^B$ statistics. For the LM-based procedures, the number of lagged difference terms in Δx_t added to the fitted regression (7) is determined using BIC selection starting from a maximum value of 6. The same number of lagged difference terms is employed for the $SupF$-based procedures in the fitted regression (10).

The entries in parentheses in the column labelled x_t are bootstrap p-values for a standard KPSS statistic applied to each predictor (with the long run variance estimate based on the quadratic spectral kernel with automatic bandwidth selection), obtained using the wild bootstrap method of Cavaliere and Taylor (2005) with 499 bootstrap replications. These p-values are small in all cases, implying rejection of the null of stationarity against the unit root alternative for each series. As is well known, the KPSS test also rejects stationarity with high probability when the series under test displays local-to-unit root behaviour, so at the very least these results are indicative of a high degree of persistence being present in each of the predictor series.

The entries in parentheses underneath the main entries are the bootstrap p-values for the LM_1^B, LM_x^B, LM_{1x}^B, $SupF_1^B$, $SupF_x^B$ and $SupF_{1x}^B$ statistics, based on $B = 499$ bootstrap replications. Considering first the results for the full sample period 1926-2015, strong evidence against H_0 being true is provided by the LM_1^B and LM_x^B tests statistics for the R_t-DY_t pairing, and, to a lesser extent, for the EP_t-DY_t pairing via LM_x^B. No evidence against H_0 is seen (i.e. no rejection at conventional significance levels) for the DE_t and LTR_t predictors, regardless of whether R_t or EP_t is employed. Turning to the pre-crisis sub-sample, evidence against H_0 is again seen for R_t-DY_t (now also including $SupF_{1x}^B$ at the 0.10-level). Some evidence is also found again for EP_t-DY_t, this time via the $SupF_{1x}^B$ test rather than the LM_x^B test. The change of sample period has no effect on the lack of rejections when using the DE_t and LTR_t predictors.

Table 2(a) also reports, under $|IV|$, the absolute value of the heteroskedasticity-robust IV t-test of Breitung and Demetrescu (2015) for predictability of y_t by x_{t-1}. This statistic combines fractional and sine function instruments and tests the significance of the estimated coefficient on x_{t-1}, having a standard normal limit distribution under the null of no predictability. Its p-value is reported in parentheses. According to $|IV|$, there is strong evidence of predictability in both the R_t-LTR_t and EP_t-LTR_t relationships when the 1926-2007 sub-sample is considered. Interestingly, neither of these pairings were found to be subject to parameter instability according to our battery of bootstrap procedures.

To informally examine the extent to which parameter instability appears present in these PRs, Figure 5 plots rolling window IV coefficient estimates and approximate 0.10-level standard error bounds. These are based on a rolling window length set at $\lfloor 0.25T \rfloor$ observations, and the x-axis dates correspond to the end of a given window sub-sample. Although it is difficult to make any firm conclusions, on examining Figure 5, we might be led to tentatively conclude that the most pronounced parameter variation is associated with the R_t-DY_t and EP_t-DY_t pairings (Figures 5(a) and 5(d)). This would be in line with our bootstrap test outcomes in Table 2(a). Also, it is credible to consider that the least pronounced parameter variation observed is associated with R_t-DE_t and EP_t-DE_t (Figures 5(b) and 5(e)), which would tie in with the generally large p-values for the associated instability tests. The estimated parameter values are also generally fairly close to zero, which is in line with $|IV|$ in Table 2(a) finding no evidence of predictability. The parameter estimates for R_t-LTR_t and EP_t-LTR_t (Figures 5(c) and 5(f)) display relative constancy at positive values over much of the sample period, which is compatible with our instability tests not rejecting, yet at the same time $|IV|$ indicating predictability for the earlier sub-sample. That the rolling parameter estimates reduce to insignificant levels towards the end of the full sample period could explain why $|IV|$ does not reject for the full sample; on the other hand, it appears from the instability test results that this change is not substantial enough, in either magnitude or duration, to be detected by our test procedures.

In Table 2(b) we consider instability tests allowing for multiple predictors, using two predictors together by combining DY_t, the predictor for which most evidence of instability was found, with either DE_t or LTR_t in the PR. For each test we use subscripts to denote the regressor coefficients permitted to vary under the alternative, with $x_1 = DY$ and $x_2 = DE$ or $x_2 = LTR$. For brevity we only show a subset of the possible statistics that could be computed (we do not report statistics that allow for variability in both the intercept and a single predictor alone). Interestingly, the pre-crisis 1926-2007 period shows little in the way of parameter instability whenever DY_t and LTR_t are combined together ($LM^B_{x_1x_2}$ being the exception). In Table 2(a), parameter instability was indicated when using DY_t alone as a predictor, but it appears that when LTR_t (which was identified as a potentially valid predictor for this period) is included in the PR, the appearance of parameter instability in the DY_t coefficient is removed, suggesting that the Table 2(a) instability results might be driven by under-specification of the PR. In the full sample, parameter instability is still detected for the DY_t and LTR_t combination; one possible explanation is the apparent late change in the LTR_t rolling coefficients observed in Figures 5(c) and 5(d), the impact of which could prevent a stable PR incorporating DY_t and LTR_t from holding for the full sample period. We also see that the addition of DE_t to the R_t-DY_t and EP_t-DY_t regressions results in no evidence for instability, despite there being evidence for DY_t coefficient instability when considered in isolation. Given that there was no evidence for DE_t being a valid predictor, a possible interpretation is that the addition of this regressor has reduced the power of the instability tests. What it is clear is that allowing multiple predictors opens the door for rather complex interactions in the parameter instability testing context.

9 Conclusions

We have developed asymptotically valid tests for structural change in the slope and/or intercept parameters of a PR model, based on the well known $SupF$ and Cramer-von-Mises type structural instability test statistics of Andrews (1993) and Nyblom (1989), respectively. To allow for an unknown degree of persistence in the predictors, and for both conditional and unconditional heteroskedasticity, a fixed regressor wild bootstrap test procedure was proposed and its asymptotic validity established. Our validity argument involved demonstrating that the asymptotic distributions of the bootstrap parameter constancy statistics, conditional on the data, coincide with the asymptotic null distributions of the corresponding statistics computed on the data, conditional on the predictors. In doing so we have shown that the standard approach to asymptotic bootstrap validity, based on bootstrap consistency for the unconditional limiting distributions of the original test statistics, is not generally applicable in cases where the bootstrap procedure treats non-stationary regressors as fixed. Monte Carlo simulations were reported which suggested that our proposed methods work well. An empirical illustration using well-known U.S. stock market data highlighted the potential value of our procedure in practice.

References

Andrews, D.W.K. (1993). Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. Econometrica 66, 47-78.

Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. Journal of Applied Econometrics 18, 1-22.

Boswijk P., G. Cavaliere, A. Rahbek and A.M.R. Taylor (2016). Inference on co-integration parameters in heteroskedastic vector autoregressions. Journal of Econometrics 192, 64–85.

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

Cai, Z., Y. Wang and Y. Wang (2015). Testing instability in a predictive regression model with nonstationary regressors. Econometric Theory 31, 953–980.

Campbell, J.Y. and M. Yogo (2006). Efficient tests of stock return predictability. Journal of Financial Econometrics 81, 27–60.

Cavaliere, G. and A.M.R. Taylor (2005). Stationarity tests under time-varying second moments. Econometric Theory 21, 1112–1129.

Cavaliere, G. and A.M.R. Taylor (2006). Testing the null of co-integration under permanent variance shifts. Journal of Time Series Analysis 27, 613–636.

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

Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. Journal of Financial Economics 106, 157–181.

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

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.

Georgiev, I., D.I. Harvey, S.J. Leybourne and A.M.R. Taylor (2018). A bootstrap stationarity test for predictive regression invalidity. Forthcoming in Journal of Business and Economic Statistics. Also available as Essex Finance Centre Working Papers, Number 21006. Downloadable from `http://repository.essex.ac.uk/21006/`

Hansen, B.E. (1992a). Tests for parameter instability in regressions with I(1) processes. Journal of Business and Economic Statistics 10, 321–335.

Hansen, B.E. (1992b). Testing for parameter instability in linear models. Journal of Policy Modelling 14, 517–533.

Hansen, B.E. (2000). Testing for structural change in conditional models. Journal of Econometrics 97, 93–115.

Henkel, S.J., J.S. Martin and F. Nardari (2011). Time-varying short-horizon predictability. Journal of Financial Economics 99, 560–580.

Jansson, M. and M.J. Moreira (2006). Optimal inference in regression models with nearly integrated regressors. Econometrica 74, 681–714.

Johannes, M., A. Korteweg and N. Polson (2014). Sequential learning, predictability, and optimal portfolio returns. The Journal of Finance 69, 611-644.

Kostakis, A., T. Magdalinos and M.P. Stamatogiannis (2015). Robust econometric inference for stock return predictability. Review of Financial Studies 28, 1506–1553.

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.

Linde, W. (1989). Gaussian measure of translated balls in a Banach space. Theory of Probability and its Applications 34, 349–359.

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.

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

Shin, Y. (1994). A residual-based test of the null of cointegration against the alternative of no cointegration. Econometric Theory 10, 91–115.

Timmermann, A. (2008). Elusive return predictability. International Journal of Forecasting 24, 1-18.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455–1508.

White, H. (1982). Maximum likelihood estimation of misspecified models, Econometrica 50, 1–25.

A Appendix

Preliminaries: In what follows we set $s_{x,0} = 0$ without loss of generality. We also define the centred variables $\mathring{y}_t := y_t - \bar{y}$, $\mathring{x}_t := x_t - \bar{x}_{-1}$ 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 $\overline{\Delta x} = T^{-1}\sum_{t=1}^{T} \Delta x_t$.

By Lemma A.1 of Boswijk *et al.* (2016), we observe that under Assumption 1,

$$T^{-1}\sum_{t=1}^{T}\epsilon_t\epsilon_t' \xrightarrow{p} H\left[\int_0^1 diag\{d_1^2(r),d_2^2(r),d_3^2(r),d_4^2(r)\}dr\right]H', \tag{A.1}$$

where $diag\{v\}$ denotes a diagonal matrix with v on the main diagonal. Next, by a routine argument, (11) implies the following OU convergence result,

$$T^{-1/2}\begin{bmatrix} x_{\lfloor T\cdot\rfloor} \\ s_{\alpha\lfloor T\cdot\rfloor} \\ s_{\beta\lfloor T\cdot\rfloor}\end{bmatrix} \xrightarrow{w} \int_0^{\cdot}\begin{bmatrix} e^{-(\cdot-s)c_x}dM_{\eta x}(s) \\ e^{-(\cdot-s)c_\alpha}dM_{\eta\alpha}(s) \\ e^{-(\cdot-s)c_\beta}dM_{\eta\beta}(s)\end{bmatrix} = \begin{bmatrix} M_{\eta x,c_x}(\cdot) \\ M_{\eta\alpha,c_\alpha}(\cdot) \\ M_{\eta\beta,c_\beta}(\cdot)\end{bmatrix} =: M_{\eta c}(\cdot) \tag{A.2}$$

and the associated convergence to stochastic integrals

$$T^{-1}\sum_{t=1}^{T}\begin{bmatrix} x_{t-1} \\ s_{\alpha,t-1} \\ s_{\beta,t-1}\end{bmatrix}[\epsilon_t',\Delta x_t,\Delta s_{\alpha t},\Delta s_{\beta t}] \xrightarrow{w} \int_0^1 M_{\eta c}(s)d[M_\eta(s)',M_{\eta c}(s)']. \tag{A.3}$$

Proof of Theorem 1: In what follows we may set $\alpha=\mu=0$ and $\beta=-T^{-1}c_xh_{21}$, without loss of generality, noting that the statistics of our interest depend on y_t only through the residuals $\hat{e}_t$ and $\hat{e}_t(\tau)$ which are invariant to the parameters α,μ and β. We further define $y_t^x := y_t - h_{21}\Delta x_t$, $\mathring{y}_t^x := \mathring{y}_t - h_{21}\Delta\mathring{x}_t$ and $\epsilon_{yt}^x := \epsilon_{yt} - h_{21}d_{1t}e_{1t} = d_{2t}e_{2t}$.

Corresponding to (6) and using invariance with respect to nonsingular linear transformations of $\mathbf{a}_t$, we have that $LM_{1x} := \frac{1}{T\hat{\sigma}^2}\sum_{i=1}^{T}\hat{S}_i'V_T^{-1}\hat{S}_i$ with $\hat{S}_i := \frac{1}{T^{1/2}}\sum_{t=1}^{i}D_T\mathring{\mathbf{a}}_{t-1}\hat{e}_t = \frac{1}{T^{1/2}}\sum_{t=1}^{i}[1,T^{-1/2}\mathring{x}_{t-1}]'\hat{e}_t$, and $V_i := \frac{1}{T}\sum_{t=1}^{i}D_T\mathring{\mathbf{a}}_{t-1}\mathring{\mathbf{a}}_{t-1}'D_T = \frac{1}{T}\sum_{t=1}^{i}[1,T^{-1/2}\mathring{x}_{t-1}]'[1,T^{-1/2}\mathring{x}_{t-1}]$, where $D_T := diag\{1,T^{-1/2}\}$ is a normalisation matrix and $\mathring{\mathbf{a}}_t := [1,\mathring{x}_t]'$. Additionally, it is seen by means of a standard argument that the process of F-statistics indexed by the break fraction r simplifies to

$$F(r) = \frac{\hat{S}_{\lfloor Tr\rfloor}'(V_{\lfloor Tr\rfloor}-V_{\lfloor Tr\rfloor}V_T^{-1}V_{\lfloor Tr\rfloor})^{-1}\hat{S}_{\lfloor Tr\rfloor}}{\hat{\sigma}^2(r)} + o_p(1) \tag{A.4}$$

uniformly over $r\in\Lambda$. The weak limit of $V_{\lfloor Tr\rfloor}$ is straightforwardly obtained using applications of the continuous mapping theorem (CMT); in particular, $D_T\mathring{\mathbf{a}}_{\lfloor Tr\rfloor} \xrightarrow{w} [1,\bar{M}_{\eta x,c_x}(r)]' := \mathbf{A}(r)$ and, hence, $V_{\lfloor Tr\rfloor} \xrightarrow{w} \int_0^r \mathbf{A}(s)\mathbf{A}'(s)\,ds =: \mathbf{V}(r)$.

In contrast, establishing the weak limit of $\hat{S}_{\lfloor Tr\rfloor}$ requires some preliminary work. To that end, consider first the limit of the partial sum process for $\hat{e}_t$, which we write as

$$T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\hat{e}_t = T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\mathring{y}_t - \left[\begin{array}{cc} T^{-3/2}\sum_{t=1}^{\lfloor Tr\rfloor}\mathring{x}_{t-1} & T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\Delta\mathring{x}_t\end{array}\right]N_T\hat{\boldsymbol{\beta}} \tag{A.5}$$

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

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

Setting $\beta_z = T^{-1}$ and $z_t = T^{-p_\alpha} g_\alpha s_{\alpha,t+1} + T^{-p_\beta} g_\beta x_{t1} s_{\beta,t+1}$, where $p_\alpha = 0$ and $p_\beta = -1/2$ for the stochastic specification **S**, and $p_\alpha = -1/2$ and $p_\beta = -1$ for the non-stochastic specification **N**, we can therefore write

$$y_t = \beta x_{t-1} + \beta_z z_{t-1} + \epsilon_{yt}, \qquad t = 1, ..., T, \tag{A.6}$$

as in eq. (1) of GHLT. Here $T^{-1/2} z_{\lfloor Tr \rfloor} \xrightarrow{w} g_\alpha M_{\eta\alpha,c_\alpha}(r) + g_\beta M_{\eta x,c_x}(r) M_{\eta\beta,c_\beta}(r)$ for the stochastic specification **S** and $T^{-1/2} z_{\lfloor Tr \rfloor} \xrightarrow{w} \{g_\alpha + g_\beta M_{\eta x,c_x}(r)\} 1(r \geq \tau_0)$ for the non-stochastic specification **N**, in $\mathcal{D}$ as $T \to \infty$. In either case, we denote the weak limit by $Q(r)$ and the corresponding de-meaned process by $\bar{Q}(r) := Q(r) - \int_0^1 Q(s)\, ds$. Since $\beta_z = O(T^{-1})$, as in GHLT, and $T^{-1/2} z_{\lfloor Tr \rfloor}$ converges weakly in $\mathcal{D}$, also as in GHLT (albeit to a different limit), by the same argument as is used in the proof of Theorem 2 in GHLT (which is based on orders of magnitude and not the exact distribution of the weak limit of $T^{-1/2} z_{\lfloor Tr \rfloor}$), we can conclude that

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{e}_t = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \mathring{y}_t^x - \frac{\sum_{t=1}^T \mathring{x}_{t-1} y_t^x}{T^{-1} \sum_{t=1}^T \mathring{x}_{t-1}^2} T^{-3/2} \sum_{t=1}^{\lfloor Tr \rfloor} \mathring{x}_{t-1} + \rho_T(r), \tag{A.7}$$

where $\sup_{r \in [0,1]} |\rho_T(r)| = o_p(1)$. Furthermore,

$$\begin{aligned} T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \mathring{y}_t^x &= T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \epsilon_{yt}^x + T^{-1/2} \beta_z \sum_{t=1}^{\lfloor Tr \rfloor} z_{t-1} - \frac{\lfloor Tr \rfloor - 1}{T^{3/2}} \left\{ \sum_{t=1}^T \epsilon_{yt}^x + \beta_z \sum_{t=1}^T z_{t-1} \right\} \\ &\xrightarrow{w} \left\{ \int_0^1 d_2^2(s) \right\}^{1/2} \{B_{\eta 2}(r) - r B_{\eta 2}(1)\} + \int_0^r \bar{Q}(s) ds =: Z(r) \end{aligned} \tag{A.8}$$

in $\mathcal{D}$, whereas

$$\begin{aligned} T^{-1} \sum_{t=1}^T \mathring{x}_{t-1} y_t^x &= T^{-1} \sum_{t=1}^T \mathring{x}_{t-1} \epsilon_{yt}^x + T^{-1} \beta_z \sum_{t=1}^T \mathring{x}_{t-1} z_{t-1} \\ &\xrightarrow{w} \left\{ \int_0^1 d_2^2(s) \right\}^{1/2} \int_0^1 \bar{M}_{\eta x,c_x}(s) dB_{\eta 2}(s) + \int_0^1 \bar{M}_{\eta x,c_x}(s) Q(s) ds \\ &= \int_0^1 \bar{M}_{\eta x,c_x}(s) dZ(s) \end{aligned} \tag{A.9}$$

using (A.2), (A.3) and applications of the CMT. Finally, using these results, we obtain the weak limits of $T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{e}_t$ and $T^{-1} \sum_{t=1}^{\lfloor Tr \rfloor} \mathring{x}_{t-1} \hat{e}_t$ as

$$\begin{aligned} \hat{S}_{\lfloor Tr \rfloor} &= T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} D_T \mathring{\mathbf{a}}_{t-1} \hat{e}_t \xrightarrow{w} \int_0^r \mathbf{A}(s)\, dH(s) \\ &= \int_0^r \mathbf{A}(s)\, dZ(s) - \mathbf{V}(r) \{\mathbf{V}(1)\}^{-1} \int_0^1 \mathbf{A}(s)\, dZ(s) \\ &= \left\{ \int_0^1 d_2^2(s) \right\}^{1/2} \left[\int_0^r \mathbf{A}(s)\, dY(s) - \mathbf{V}(r) \{\mathbf{V}(1)\}^{-1} \int_0^1 \mathbf{A}(s)\, dY(s) \right] \end{aligned} \tag{A.10}$$

in $\mathcal{D}_2$, where $H(r) := Z(r) - \left\{ \int_0^1 \bar{M}_{\eta x,c_x}^2(s) \right\}^{-1} \int_0^1 \bar{M}_{\eta x,c_x}(s) dZ(s) \int_0^r \bar{M}_{\eta x,c_x}(s)$ is the integrated residual of a continuous-time least squares regression of $dY(r)$ on $\bar{M}_{\eta x,c_x}(r) dr$.

Regarding the variance estimators used in constructing the statistics, using an orders of magnitude based argument as in the proof of Theorem 2 in GLHT, we obtain that

$$\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}(\epsilon_{yt}^{x})^2 + o_p(1) \xrightarrow{p} \int_0^1 d_2(r)^2. \quad \text{(A.11)}$$

On the other hand, from the definition (9) of $F(r)$ and (A.4), it follows that $\hat{\sigma}^2(r) = \hat{\sigma}^2 - T^{-1}\hat{S}'_{\lfloor Tr\rfloor}(V_{\lfloor Tr\rfloor} - V_{\lfloor Tr\rfloor}V_T^{-1}V_{\lfloor Tr\rfloor})^{-1}\hat{S}_{\lfloor Tr\rfloor} + o_p(T^{-1})$, uniformly in $r \in \Lambda$.

By combining the previous results and using applications of the CMT, we obtain that

$$LM_{1x} \xrightarrow{w} \frac{1}{\int_0^1 d_2^2(r)} \int_0^1 \left(\int_0^r \mathbf{A}'(s)\, dH(s) \{\mathbf{V}(1)\}^{-1} \int_0^r \mathbf{A}(s)\, dH(s) \right) dr$$

and

$$SupF_{1x} \xrightarrow{w} \frac{1}{\int_0^1 d_2^2(r)} \sup_{r\in\Lambda} \left(\int_0^r \mathbf{A}'(s)\, dH(s) \{\mathbf{V}(r) - \mathbf{V}(r)\mathbf{V}(1)^{-1}\mathbf{V}(r)\}^{-1} \int_0^r \mathbf{A}(s)\, dH(s) \right)$$

which reduce to the stated expressions in Theorem 1 on normalising H by $\{\int_0^1 d_2(r)^2\}^{1/2}$. ■

Before progressing to the proof of Theorem 2 we define the conditional convergence modes which will be used in the rest of the Appendix. Let ξ_T (respectively, η_T) be random elements of a Polish space, defined on the same probability space as the original data (respectively, the original and the bootstrap data), and ξ, η be random elements of a Polish space defined on the same probability space as B_1. For weak convergence of random measures induced by conditioning, i.e., of the form $\xi_T|x \xrightarrow{w} \xi|B_1$ and $\eta_T|x,y \xrightarrow{w} \eta|B_1$, we write resp. $\xi_T \xrightarrow{w_x} \xi|B_1$ and $\eta_T \xrightarrow{w^*} \eta|B_1$, the definitions being $E\{f(\xi_T)|x\} \xrightarrow{w} E\{f(\xi)|B_1\}$ and $E\{g(\eta_T)|x,y\} \xrightarrow{w} E\{g(\eta)|B_1\}$ for all bounded continuous real functions f and g with matching domain. Importantly, we say that the w_x and w^* convergence are *joint* if $(E\{f(\xi_T)|x\}, E\{g(\eta_T)|x,y\})' \xrightarrow{w} (E\{f(\xi)|B_1\}, E\{g(\eta)|B_1\})'$ for the same class of functions f, g. This is the meaning of joint convergence in Theorems 2 and A.1. We notice that it is distinct from two w_x convergence results $\xi'_T \xrightarrow{w_x} \xi'|B_1$ and $\xi''_T \xrightarrow{w_x} \xi''|B_1$ being joint, or equivalently, from $\xi_T \xrightarrow{w_x} \xi|B_1$ with $\xi_T = (\xi'_T, \xi''_T)$ and $\xi = (\xi', \xi'')$, where $E\{f(\xi'_T, \xi''_T)|x\} \xrightarrow{w} E\{f(\xi', \xi'')|B_1\}$ should hold for bounded continuous f (and similarly, for w^*).

We next report in Theorem A.1 some results from GHLT, adapted to the problem discussed here and which will subsequently be used in the proof of Theorem 2.

Theorem A.1. *Let $\tilde{e}_{Tt}$ $(t = 1, ..., T)$ be scalar measurable functions of the data x, y and such that $T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\tilde{e}_{Tt}^2 \xrightarrow{p} \int_0^r m^2(s)ds$ for $r \in [0,1]$, where $m(\cdot)$ is a square-integrable real function on $[0,1]$. Introduce $\epsilon_{tb}^* := w_t\tilde{e}_{Tt}$ $(t = 1, ..., T)$, and $B_m^*(\cdot) := \int_0^{\cdot} m(s)dB^*(s)$, where B^* is a standard Brownian motion independent of B (and thus, of M_η) of (11). Under Assumption 2, the following converge jointly as $T \to \infty$:*

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

in the sense of weak convergence of random measures on $\mathcal{D}^7$*, and*

$$\left(T^{-1/2} \sum_{t=1}^{\lfloor T\cdot \rfloor} e_{1t}, T^{-1/2} \sum_{t=1}^{\lfloor T\cdot \rfloor} \epsilon^*_{tb}, T^{-1} \sum_{t=1}^{\lfloor T\cdot \rfloor} \sum_{s=1}^{t-1} \epsilon_{xs} \epsilon^*_{tb} \right) \overset{w^*}{\to} \left(B_1(\cdot), B^*_m(\cdot), \int_0^{\cdot} M_{\eta x}(s) dB^*_m(s) \right) \Big| B_1$$

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

Similarly to the discussion of Theorem 3 in GHLT, Theorem A.1 implies that $T^{-1/2} x_{\lfloor T\cdot \rfloor} \overset{w_x}{\to} M_{\eta x, c_x}(\cdot)|B_1$, $T^{-1/2} s_{\alpha \lfloor T\cdot \rfloor} \overset{w_x}{\to} M_{\eta\alpha, c_\alpha}(\cdot)|B_1$ and $T^{-1/2} s_{\beta \lfloor T\cdot \rfloor} \overset{w_x}{\to} M_{\eta\beta, c_\beta}(\cdot)|B_1$, jointly with the convergence in Theorem A.1. Furthermore, regarding stochastic integrals,

$$T^{-1} \sum_{t=1}^{T} s_{x,t-1} \epsilon_{it} \overset{w_x}{\to} \int_0^1 M_{\eta x, c_x}(s) dM_{\eta i}(s) \Big| B_1, \; i \in \{y, \alpha, \beta\},$$

jointly with the convergence in Theorem A.1 and its implications. By the CMT, as $T^{-2} \sum_{t=1}^{T} s_{x,t-1} z_{t-1} \overset{w_x}{\to} \int_0^1 M_{\eta x, c_x}(s) \, Q(s) \, ds | B_1$ 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 $\epsilon^x_{yt} := \epsilon_{yt} - h_{21} d_{1t} e_{1t} = d_{2t} e_{2t}$ that

$$\begin{aligned} T^{-1} \sum_{t=1}^{T} \mathring{s}_{x,t-1} y^x_t &= T^{-1} \sum_{t=1}^{T} \mathring{s}_{x,t-1} (\epsilon^x_{yt} + \beta_z z_{t-1}) \qquad \text{(A.12)} \\ &\overset{w_x}{\to} \left\{ \left\{ \int_0^1 d_2^2(s) \right\}^{1/2} \int_0^1 \bar{M}_{\eta x, c_x}(s) \, dB_{\eta 2}(s) + \int_0^1 \bar{M}_{\eta x, c_x}(s) \, Q(s) \right\} \Big| B_1, \end{aligned}$$

if $\beta = -T^{-1} c_x h_{21}$ in (1) and (A.6).

Proof of Theorem 2: We may again set $\alpha = \mu = 0$ and $\beta = -T^{-1} c_x h_{21}$ without loss of generality. For the original statistics LM_{1x} and $SupF_{1x}$, we argue that the proof given for Theorem 1 also remains valid conditionally. To that end, notice first that for a general sequence ξ_T of random elements defined on the probability space of the data, if $\xi_T \overset{p}{\to} K$, where K is a constant, then it follows that $E_x f(\xi_T) \overset{p}{\to} f(K)$ for bounded continuous real f, so equivalently,

$$\xi_T \overset{w_x}{\to} K. \qquad \text{(A.13)}$$

Using (A.13), the remainder term in (A.7) is then seen to satisfy $\sup_{r \in [0,1]} |\rho_T(r)| \overset{w_x}{\to} 0$. Moreover, corresponding to (A.8) and (A.9), we have that $\sum_{t=1}^{\lfloor Tr \rfloor} \mathring{y}^x_t \overset{w_x}{\to} Z(r) | B_1$ and $T^{-1} \sum_{t=1}^{T} \mathring{x}_{t-1} y^x_t \overset{w_x}{\to} \int_0^1 \bar{M}_{\eta x, c_x}(s) dZ(s) \Big| B_1$, by Theorem A.1 and its discussion (see (A.12)), and the CMT. Since $D_T \mathring{\mathbf{a}}_{\lfloor Tr \rfloor} \overset{w_x}{\to} \mathbf{A}(r) | B_1$ and $V_{\lfloor Tr \rfloor} \overset{w_x}{\to} \mathbf{V}(r) | B_1$, jointly in $\mathcal{D}_6$ as a direct consequence of the analogous unconditional convergence and the CMT (because the right-hand sides are measurable with respect to x and the left-hand sides with respect to B_1), by combining the previous results we find that $\hat{S}_{\lfloor Tr \rfloor} \overset{w_x}{\to} \int_0^r \mathbf{A}(s) \, dH(s) \Big| B_1$ in $\mathcal{D}_2$ as the conditional counterpart of (A.10). Again using (A.13) it follows that the expansion in (A.4) holds also conditionally on x. Next, (A.11) and (A.13) with $\xi_T = \hat{\sigma}^2$ imply that $\hat{\sigma}^2 \overset{w_x}{\to} \int_0^1 d_2(r)^2$. Using the conditional convergence of $V_{\lfloor Tr \rfloor}$ and $\hat{S}_{\lfloor Tr \rfloor}$, we obtain the results given in Theorem 2 concerning the original LM_{1x} and $SupF_{1x}$ statistics.

Turning next to the bootstrap LM_{1x}^* and $SupF_{1x}^*$ statistics, we observe first that a bootstrap invariance principle holds jointly with the previously stated convergence results. The bootstrap partial-sum process $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} y_t^*$ is of the form of $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\epsilon_{tb}^*$ of Theorem A.1, 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.* (2016), we conclude that $T^{-1}\sum_{t=1}^{\lfloor Tr\rfloor}\hat{e}_t^2 \xrightarrow{p} \int_0^r d_2^2(s)ds = \int_0^r m^2(s)ds$ with $m = d_2$. For this choice of m, from Theorem A.1 and its discussion it follows that

$$\left(T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} e_{1t}, T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} y_t^*, T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}\sum_{s=1}^{t-1}\epsilon_{xs}y_t^*\right) \xrightarrow{w^*} \left(B_1, B_m^*, \int_0^{\cdot} M_{\eta x}(s)dB_m^*(s)\right)\Big| B_1 \quad \text{(A.14)}$$

jointly with $T^{-1/2}x_{\lfloor T\cdot\rfloor} \xrightarrow{w^*} M_{\eta x,c_x}|B_1$ and the conditional convergence of LM_{1x} and $SupF_{1x}$.

Next, with $\hat{\epsilon}_{yt}^*$ denoting the residuals from the bootstrap analogue of the regression in (7), it holds that

$$\begin{aligned} S_{\lfloor Tr\rfloor}^* := \quad & T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor} D_T\mathring{\mathbf{a}}_{t-1}\hat{\epsilon}_{yt}^* = \sum_{t=1}^{\lfloor Tr\rfloor}\begin{bmatrix} T^{-1/2} \\ T^{-1}\mathring{x}_{t-1}\end{bmatrix}(y_t^* - \bar{y}^*) \\ & -\frac{T^{-1}\sum_{t=1}^T \mathring{x}_{t-1}y_t^*}{T^{-2}\sum_{t=1}^T \mathring{x}_{t-1}^2}\sum_{t=1}^{\lfloor Tr\rfloor}\begin{bmatrix} T^{-3/2} \\ T^{-2}\mathring{x}_{t-1}\end{bmatrix}\mathring{x}_{t-1}, \end{aligned}$$

where, by (A.14) and the CMT, the following can be seen to converge jointly, and jointly with LM_{1x} and $SupF_{1x}$: $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}(y_t^* - \bar{y}^*) \xrightarrow{w^*} \{B_m^*(\cdot) - (\cdot)B_m^*(1)\}|B_1$, $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\mathring{x}_{t-1}(y_t^* - \bar{y}^*) \xrightarrow{w^*} \int_0^{\cdot}\bar{M}_{\eta x,c_x}(s)dB_m^*(s)|B_1$, $T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}(T^{-1/2}\mathring{x}_{t-1})^i \xrightarrow{w^*} \int_0^{\cdot}\bar{M}_{\eta x,c_x}^i(s)ds|B_1$ $(i = 1,2)$ and $T^{-1}\sum_{t=1}^T \mathring{x}_{t-1}y_t^* \xrightarrow{w^*} \int_0^1 \bar{M}_{\eta x,c_x}(s)dB_m^*(s)|B_1$ analogously to (A.12). Since the limit processes in $\mathcal{D}$ are continuous, it further holds that

$$\begin{aligned} & S_{\lfloor Tr\rfloor}^* \xrightarrow{w^*} \{\int_0^1 d_2^2(s)\}^{1/2}\mathbf{J}_0^*(r)\Big| B_1 := \int_0^r \mathbf{A}(s)\,dH^*(s)\Big| B_1 \\ = \quad & \int_0^r \mathbf{A}(s)\,d\{B_m^*(s) - sB_m^*(1)\} - \mathbf{V}(r)\{\mathbf{V}(1)\}^{-1}\int_0^1 \mathbf{A}(s)\,d\{B_m^*(s) - sB_m^*(1)\}\Big| B_1 \\ = \quad & \int_0^r \mathbf{A}(s)\,dB_m^*(s) - \mathbf{V}(r)\{\mathbf{V}(1)\}^{-1}\int_0^1 \mathbf{A}(s)\,dB_m^*(s)\Big| B_1 \end{aligned}$$

in $\mathcal{D}_2$, jointly with LM_{1x} and $SupF_{1x}$, and where

$$H^*(r) = B_m^*(r) - rB_m^*(1) - \left\{\int_0^1 \bar{B}_{1\eta,c_x}^2(s)\right\}^{-1}\int_0^1 \bar{B}_{1\eta,c_x}(s)dB_m^*(s)\int_0^r \bar{B}_{1\eta,c_x}(s).$$

It can then be directly checked that $(\mathbf{J}_0^*, B_1) \stackrel{d}{=} (\mathbf{J}_0, B_1)$, so $E(f(\mathbf{J}_0^*, B_1)|B_1) = E(f(\mathbf{J}_0, B_1)|B_1)$ a.s. for every continuous real function f with conformable domain. This allows us to derive the limits in Theorem 2 with $\mathbf{J}_0$ in place of $\mathbf{J}_0^*$.

Finally, using the foregoing convergence results, the residual variance from the fitted bootstrap regression analogue of (7) can be seen to satisfy

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

because $E\{T^{-1}\sum_{t=1}^{T}(w_t^2-1)\hat{e}_t^2|x,y\}^2 = 2T^{-2}\sum_{t=1}^{T}\hat{e}_t^4 = o_p(1)$ under the assumption of finite fourth moments; here $o_p^*(1)$ denotes terms such that $o_p^*(1) \overset{w^*}{\to} 0$. We conclude that $\hat{\sigma}^{*2} \overset{w^*}{\to} \int_0^1 d_2^2(s)$ and, by the CMT, from

$$LM_{1x}^* = \frac{1}{T\hat{\sigma}^{*2}}\sum_{i=1}^{T} S_i^{*\prime}V_T^{-1}S_i^* \text{ and } F^*(r) = \frac{S_{\lfloor Tr\rfloor}^{*\prime}(V_{\lfloor Tr\rfloor} - V_{\lfloor Tr\rfloor}V_T^{-1}V_{\lfloor Tr\rfloor})^{-1}S_{\lfloor Tr\rfloor}^*}{\hat{\sigma}^{*2}(r)}$$

with $\hat{\sigma}^{*2}(r) = \hat{\sigma}^{*2} - T^{-1}S_{\lfloor Tr\rfloor}^{*\prime}(V_{\lfloor Tr\rfloor} - V_{\lfloor Tr\rfloor}V_T^{-1}V_{\lfloor Tr\rfloor})^{-1}S_{\lfloor Tr\rfloor}^*$ $(r \in \Lambda)$, it follows that LM_{1x}^* and $SupF_{1x}^*$ converge as asserted, jointly with LM_{1x} and $SupF_{1x}$. □

Proof of Corollary 1.

The random cdf's conditional on B_1 of the conditional limit distributions given in Theorem 2 are continuous a.s. For the LM_{1x} statistic this follows from the representation of the limit distribution conditional on B_1 as the distribution of an infinite weighted sum of independent $\chi^2(2)$ variables, similarly to Nyblom (1989) and Rao and Swift (2006, pp.472-473), using the continuity of $\mathbf{V}$ a.s. For $SupF_{1x}$ continuity of the limiting conditional cdf follows from Proposition 3.2 of Linde (1989) applied conditionally on B_1. The proof then proceeds as that of Corollary 1 in GHLT. □

Table 1 (a). Finite sample size of tests, and power of tests under H^S.

τ_{0h}	σ	g	LM_1^B	LM_x^B	LM_{1x}^B	SF_1^B	SF_x^B	SF_{1x}^B	LM_1^B	LM_x^B	LM_{1x}^B	SF_1^B	SF_x^B	SF_{1x}^B
						Panel A. H_1^S								
			$c_\alpha = c_\beta = 0$						$c_\alpha = c_\beta = 10$					
–	1	0	0.100	0.101	0.094	0.110	0.085	0.071	0.100	0.101	0.094	0.110	0.085	0.071
		15	0.784	0.477	0.767	0.811	0.484	0.734	0.331	0.213	0.330	0.369	0.203	0.266
		35	0.954	0.740	0.961	0.968	0.777	0.952	0.697	0.469	0.726	0.778	0.496	0.717
$\frac{1}{2}$	4	0	0.102	0.102	0.104	0.124	0.103	0.099	0.102	0.102	0.104	0.124	0.103	0.099
		15	0.354	0.154	0.331	0.292	0.145	0.219	0.132	0.112	0.130	0.140	0.108	0.117
		35	0.695	0.321	0.673	0.647	0.321	0.534	0.249	0.155	0.245	0.229	0.146	0.177
	$\frac{1}{4}$	0	0.101	0.104	0.106	0.121	0.107	0.094	0.101	0.104	0.106	0.121	0.107	0.094
		15	0.865	0.567	0.874	0.860	0.585	0.795	0.440	0.270	0.461	0.421	0.257	0.333
		35	0.972	0.803	0.977	0.979	0.846	0.972	0.801	0.578	0.834	0.848	0.627	0.803
$\frac{3}{4}$	4	0	0.095	0.117	0.107	0.130	0.137	0.141	0.095	0.117	0.107	0.130	0.137	0.141
		15	0.436	0.219	0.424	0.297	0.200	0.244	0.144	0.136	0.160	0.151	0.145	0.155
		35	0.774	0.440	0.772	0.684	0.418	0.603	0.312	0.224	0.329	0.245	0.201	0.222
	$\frac{1}{4}$	0	0.101	0.100	0.104	0.116	0.092	0.080	0.101	0.100	0.104	0.116	0.092	0.080
		15	0.847	0.526	0.840	0.846	0.568	0.779	0.416	0.227	0.405	0.414	0.229	0.303
		35	0.971	0.773	0.970	0.976	0.840	0.970	0.776	0.500	0.801	0.826	0.586	0.765
						Panel B. H_x^S								
			$c_\alpha = c_\beta = 0$						$c_\alpha = c_\beta = 10$					
–	1	15	0.489	0.771	0.746	0.569	0.767	0.717	0.218	0.372	0.338	0.260	0.360	0.305
		35	0.748	0.937	0.937	0.824	0.948	0.936	0.468	0.685	0.685	0.557	0.724	0.689
$\frac{1}{2}$	4	15	0.359	0.601	0.550	0.392	0.589	0.527	0.178	0.290	0.254	0.226	0.278	0.256
		35	0.625	0.866	0.844	0.669	0.865	0.842	0.369	0.616	0.569	0.436	0.605	0.570
	$\frac{1}{4}$	15	0.379	0.646	0.587	0.404	0.604	0.540	0.192	0.312	0.271	0.213	0.301	0.247
		35	0.669	0.891	0.872	0.706	0.891	0.870	0.404	0.629	0.589	0.443	0.633	0.591
$\frac{3}{4}$	4	15	0.325	0.647	0.551	0.325	0.551	0.461	0.144	0.276	0.222	0.186	0.268	0.220
		35	0.589	0.868	0.838	0.607	0.834	0.783	0.291	0.570	0.501	0.339	0.522	0.454
	$\frac{1}{4}$	15	0.484	0.721	0.696	0.521	0.700	0.656	0.225	0.357	0.339	0.245	0.349	0.306
		35	0.750	0.911	0.924	0.791	0.925	0.919	0.481	0.672	0.681	0.529	0.700	0.672
						Panel C. H_{1x}^S								
			$c_\alpha = c_\beta = 0$						$c_\alpha = c_\beta = 10$					
–	1	15	0.811	0.782	0.914	0.862	0.793	0.895	0.404	0.422	0.496	0.473	0.427	0.456
		35	0.946	0.919	0.990	0.971	0.935	0.990	0.754	0.733	0.863	0.827	0.773	0.858
$\frac{1}{2}$	4	15	0.497	0.616	0.639	0.487	0.597	0.585	0.203	0.299	0.275	0.246	0.293	0.267
		35	0.797	0.866	0.927	0.815	0.872	0.901	0.447	0.627	0.622	0.492	0.624	0.600
	$\frac{1}{4}$	15	0.883	0.752	0.926	0.881	0.751	0.882	0.483	0.420	0.550	0.483	0.410	0.453
		35	0.970	0.905	0.990	0.978	0.913	0.987	0.818	0.720	0.885	0.860	0.750	0.861
$\frac{3}{4}$	4	15	0.523	0.649	0.692	0.446	0.570	0.525	0.190	0.294	0.258	0.206	0.276	0.235
		35	0.825	0.868	0.938	0.804	0.840	0.879	0.428	0.595	0.605	0.419	0.542	0.512
	$\frac{1}{4}$	15	0.866	0.770	0.929	0.878	0.789	0.902	0.474	0.427	0.548	0.495	0.439	0.479
		35	0.964	0.914	0.990	0.979	0.939	0.990	0.805	0.727	0.889	0.852	0.785	0.876

Table 1 (b). Finite sample power of tests under H^N.

τ_{0h}	σ	g	LM_1^B	LM_x^B	LM_{1x}^B	SF_1^B	SF_x^B	SF_{1x}^B	LM_1^B	LM_x^B	LM_{1x}^B	SF_1^B	SF_x^B	SF_{1x}^B
Panel A. H_1^N														
			$\tau_0 = \frac{1}{2}$						$\tau_0 = \frac{3}{4}$					
−	1	15	0.638	0.223	0.554	0.568	0.193	0.403	0.455	0.216	0.407	0.473	0.186	0.315
		35	0.991	0.540	0.979	0.991	0.546	0.975	0.961	0.522	0.954	0.980	0.548	0.938
$\frac{1}{2}$	4	15	0.179	0.109	0.171	0.159	0.112	0.121	0.142	0.110	0.132	0.151	0.111	0.120
		35	0.461	0.160	0.429	0.345	0.140	0.228	0.310	0.145	0.270	0.321	0.142	0.217
	$\frac{1}{4}$	15	0.838	0.291	0.804	0.717	0.250	0.544	0.714	0.242	0.715	0.531	0.207	0.304
		35	0.998	0.656	0.996	0.997	0.690	0.992	0.998	0.572	1.000	1.000	0.726	0.998
$\frac{3}{4}$	4	15	0.221	0.134	0.211	0.150	0.145	0.151	0.147	0.122	0.139	0.166	0.139	0.157
		35	0.618	0.225	0.585	0.299	0.190	0.226	0.325	0.178	0.299	0.311	0.173	0.258
	$\frac{1}{4}$	15	0.787	0.271	0.712	0.698	0.241	0.505	0.668	0.216	0.615	0.583	0.219	0.348
		35	0.998	0.624	0.994	0.997	0.665	0.990	0.997	0.522	0.998	0.999	0.774	0.993
Panel B. H_x^N														
			$\tau_0 = \frac{1}{2}$						$\tau_0 = \frac{3}{4}$					
−	1	15	0.250	0.642	0.529	0.281	0.534	0.409	0.243	0.473	0.409	0.292	0.436	0.331
		35	0.563	0.990	0.976	0.672	0.984	0.963	0.533	0.891	0.842	0.635	0.922	0.880
$\frac{1}{2}$	4	15	0.128	0.168	0.152	0.136	0.152	0.131	0.226	0.479	0.376	0.265	0.436	0.351
		35	0.264	0.425	0.388	0.231	0.465	0.341	0.500	0.938	0.873	0.579	0.916	0.873
	$\frac{1}{4}$	15	0.163	0.289	0.224	0.169	0.246	0.180	0.132	0.131	0.140	0.130	0.113	0.100
		35	0.404	0.681	0.622	0.391	0.730	0.621	0.252	0.281	0.288	0.203	0.233	0.168
$\frac{3}{4}$	4	15	0.155	0.318	0.241	0.157	0.208	0.175	0.170	0.404	0.294	0.178	0.296	0.231
		35	0.316	0.794	0.643	0.269	0.608	0.417	0.370	0.903	0.781	0.369	0.794	0.644
	$\frac{1}{4}$	15	0.254	0.572	0.480	0.258	0.480	0.372	0.191	0.185	0.196	0.187	0.187	0.142
		35	0.571	0.962	0.943	0.615	0.957	0.926	0.445	0.391	0.469	0.444	0.539	0.461
Panel C. H_{1x}^N														
			$\tau_0 = \frac{1}{2}$						$\tau_0 = \frac{3}{4}$					
−	1	15	0.607	0.607	0.740	0.606	0.536	0.660	0.482	0.460	0.572	0.532	0.452	0.525
		35	0.906	0.916	0.996	0.928	0.904	0.994	0.795	0.762	0.899	0.848	0.776	0.943
$\frac{1}{2}$	4	15	0.208	0.184	0.225	0.174	0.167	0.158	0.249	0.474	0.389	0.305	0.436	0.377
		35	0.511	0.443	0.585	0.430	0.477	0.492	0.553	0.927	0.879	0.635	0.908	0.894
	$\frac{1}{4}$	15	0.815	0.397	0.816	0.724	0.390	0.622	0.693	0.268	0.696	0.543	0.261	0.342
		35	0.995	0.704	0.997	0.993	0.761	0.994	0.990	0.544	0.992	0.990	0.662	0.977
$\frac{3}{4}$	4	15	0.257	0.331	0.317	0.177	0.217	0.190	0.201	0.409	0.315	0.221	0.298	0.246
		35	0.617	0.761	0.808	0.470	0.605	0.574	0.501	0.877	0.828	0.486	0.776	0.714
	$\frac{1}{4}$	15	0.723	0.553	0.799	0.691	0.505	0.687	0.633	0.260	0.616	0.602	0.350	0.451
		35	0.960	0.850	0.997	0.965	0.863	0.997	0.935	0.499	0.946	0.946	0.665	0.936

Table 2 (a). Application to updated Welch and Goyal (2008) data: bivariate regressions.

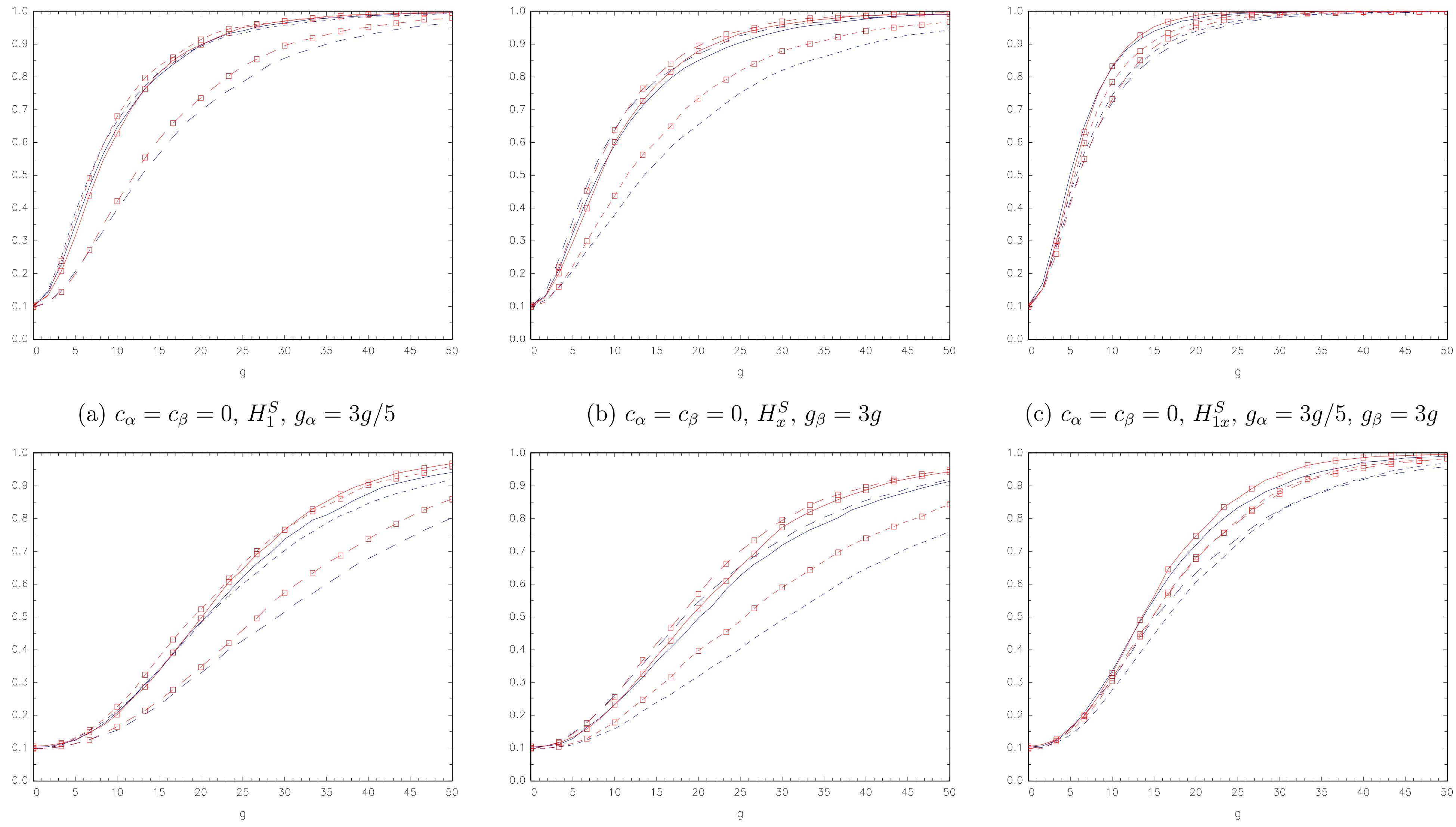
y_t	x_t	LM_1^B	LM_x^B	LM_{1x}^B	$SupF_1^B$	$SupF_x^B$	$SupF_{1x}^B$	$\lvert IV \rvert$
				Panel A. 1926-2015				
R_t	DY_t	0.326	0.326	0.364	6.969	7.054	17.47	0.742
	(0.014)	(0.026)	(0.022)	(0.104)	(0.265)	(0.271)	(0.146)	(0.458)
	DE_t	0.132	0.119	0.152	4.986	3.853	4.993	0.684
	(0.014)	(0.315)	(0.405)	(0.772)	(0.415)	(0.601)	(0.711)	(0.494)
	LTR_t	0.057	0.165	0.265	2.346	4.950	11.18	1.376
	(0.024)	(0.830)	(0.230)	(0.333)	(0.810)	(0.361)	(0.198)	(0.169)
EP_t	DY_t	0.192	0.220	0.315	9.395	9.046	16.79	1.109
		(0.148)	(0.082)	(0.182)	(0.156)	(0.172)	(0.154)	(0.268)
	DE_t	0.161	0.119	0.202	4.904	4.344	5.010	0.306
		(0.212)	(0.429)	(0.635)	(0.407)	(0.523)	(0.701)	(0.759)
	LTR_t	0.085	0.193	0.263	1.968	5.063	10.50	1.296
		(0.655)	(0.200)	(0.375)	(0.850)	(0.291)	(0.226)	(0.195)
				Panel B. 1926-2007				
R_t	DY_t	0.352	0.365	0.414	6.899	6.849	23.73	0.821
	(0.034)	(0.042)	(0.028)	(0.106)	(0.321)	(0.347)	(0.076)	(0.412)
	DE_t	0.056	0.060	0.128	4.090	3.662	4.266	1.000
	(0.008)	(0.764)	(0.719)	(0.798)	(0.579)	(0.627)	(0.794)	(0.317)
	LTR_t	0.076	0.153	0.261	2.569	5.444	12.67	2.151
	(0.018)	(0.737)	(0.301)	(0.397)	(0.794)	(0.391)	(0.186)	(0.031)
EP_t	DY_t	0.167	0.200	0.346	8.585	7.916	22.56	1.268
		(0.222)	(0.126)	(0.166)	(0.188)	(0.248)	(0.078)	(0.205)
	DE_t	0.067	0.074	0.134	3.734	2.279	3.772	0.513
		(0.657)	(0.621)	(0.752)	(0.651)	(0.860)	(0.840)	(0.608)
	LTR_t	0.190	0.248	0.349	4.395	6.761	11.77	1.947
		(0.271)	(0.168)	(0.222)	(0.477)	(0.309)	(0.226)	(0.052)

Note: Entries in parentheses are bootstrap p-values.

Table 2 (b). Application to updated Welch and Goyal (2008) data: multivariate regressions.

y_t	x_{1t}, x_{2t}	LM_1^B	$LM_{x_1}^B$	$LM_{x_2}^B$	$LM_{x_1x_2}^B$	$LM_{1x_1x_2}^B$	$SupF_1^B$	$SupF_{x_1}^B$	$SupF_{x_2}^B$	$SupF_{x_1x_2}^B$	$SupF_{1x_1x_2}^B$
						Panel A. 1926-2015					
R_t	DY_t, DE_t	0.097	0.109	0.112	0.148	0.191	5.870	5.959	6.276	6.687	15.42
		(0.311)	(0.238)	(0.349)	(0.699)	(0.768)	(0.359)	(0.353)	(0.379)	(0.561)	(0.224)
	DY_t, LTR_t	0.239	0.240	0.045	0.492	0.533	6.187	6.626	3.448	13.41	21.31
		(0.082)	(0.058)	(0.705)	(0.026)	(0.098)	(0.335)	(0.323)	(0.507)	(0.146)	(0.128)
EP_t	DY_t, DE_t	0.114	0.129	0.178	0.205	0.396	8.561	8.420	8.746	9.605	15.46
		(0.224)	(0.162)	(0.164)	(0.523)	(0.363)	(0.146)	(0.146)	(0.202)	(0.285)	(0.200)
	DY_t, LTR_t	0.143	0.163	0.074	0.398	0.536	9.220	8.153	4.336	12.05	20.11
		(0.246)	(0.166)	(0.511)	(0.068)	(0.098)	(0.152)	(0.222)	(0.417)	(0.184)	(0.148)
						Panel B. 1926-2007					
R_t	DY_t, DE_t	0.097	0.114	0.131	0.176	0.237	5.638	5.520	6.530	6.544	22.53
		(0.309)	(0.220)	(0.232)	(0.571)	(0.659)	(0.373)	(0.383)	(0.267)	(0.501)	(0.048)
	DY_t, LTR_t	0.195	0.205	0.072	0.439	0.518	6.348	6.727	3.977	15.13	24.37
		(0.170)	(0.130)	(0.517)	(0.066)	(0.150)	(0.383)	(0.365)	(0.517)	(0.126)	(0.118)
EP_t	DY_t, DE_t	0.095	0.111	0.152	0.209	0.428	7.164	7.103	7.966	8.168	22.66
		(0.327)	(0.230)	(0.170)	(0.483)	(0.281)	(0.214)	(0.232)	(0.166)	(0.327)	(0.044)
	DY_t, LTR_t	0.112	0.122	0.134	0.339	0.580	8.044	7.114	5.001	13.58	23.00
		(0.365)	(0.313)	(0.271)	(0.156)	(0.106)	(0.236)	(0.317)	(0.429)	(0.158)	(0.116)

Note: Entries in parentheses are bootstrap p-values.



(a) $c_\alpha = c_\beta = 0$, H_1^S, $g_\alpha = 3g/5$

(b) $c_\alpha = c_\beta = 0$, H_x^S, $g_\beta = 3g$

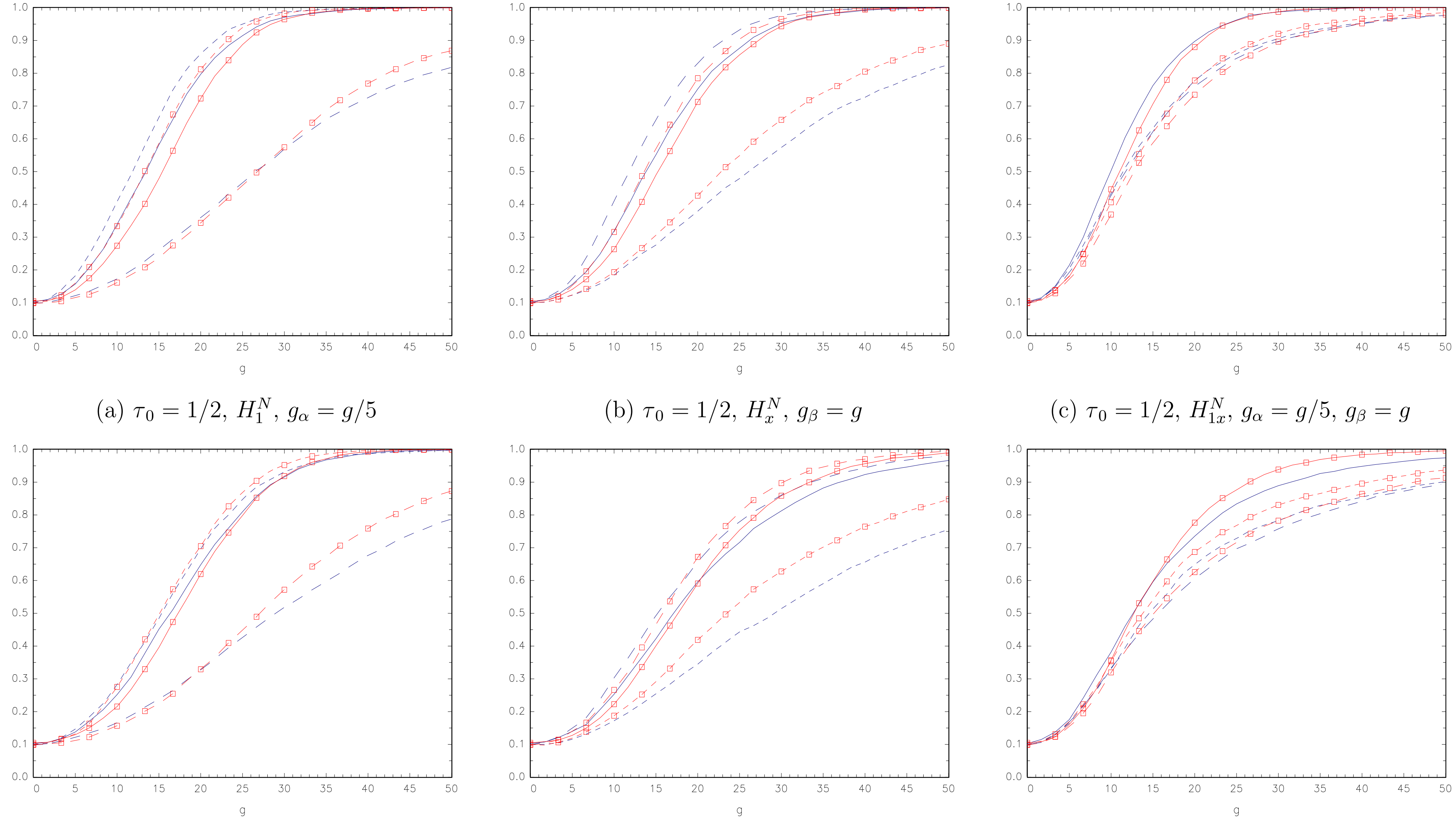
(c) $c_\alpha = c_\beta = 0$, H_{1x}^S, $g_\alpha = 3g/5$, $g_\beta = 3g$

(d) $c_\alpha = c_\beta = 10$, H_1^S, $g_\alpha = 3g/5$

(e) $c_\alpha = c_\beta = 10$, H_x^S, $g_\beta = 3g$

(f) $c_\alpha = c_\beta = 10$, H_{1x}^S, $g_\alpha = 3g/5$, $g_\beta = 3g$

Figure 1. Asymptotic local power of tests under H^S: LM_1^B: - - - , LM_x^B: — — , LM_{1x}^B: —— , $SupF_1^B$: -□-, $SupF_x^B$: —□—, $SupF_{1x}^B$: —□—



(a) $\tau_0 = 1/2$, H_1^N, $g_\alpha = g/5$

(b) $\tau_0 = 1/2$, H_x^N, $g_\beta = g$

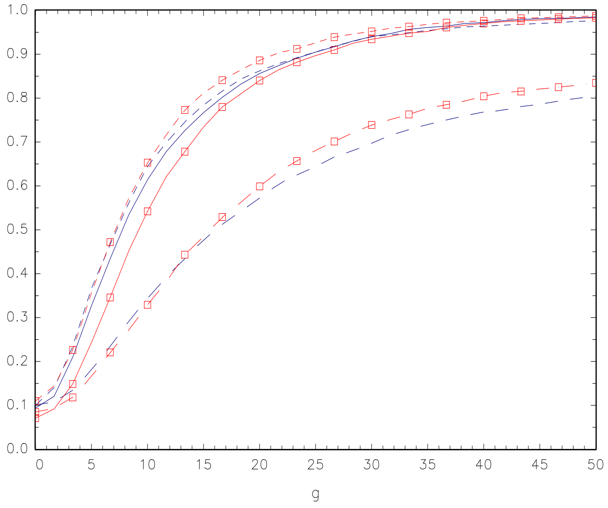
(c) $\tau_0 = 1/2$, H_{1x}^N, $g_\alpha = g/5$, $g_\beta = g$

(d) $\tau_0 = 3/4$, H_1^N, $g_\alpha = g/5$

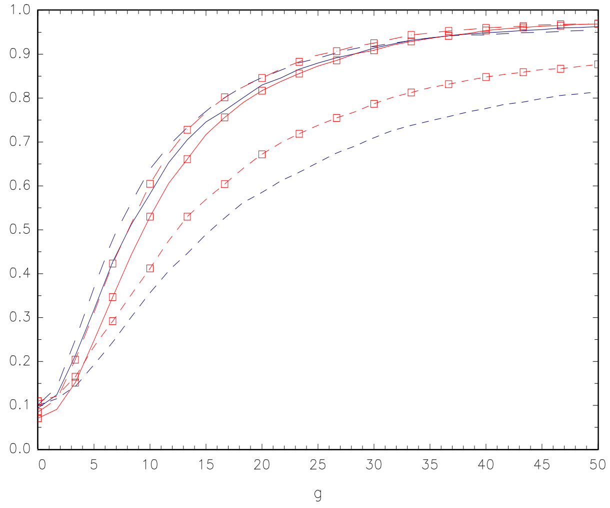
(e) $\tau_0 = 3/4$, H_x^N, $g_\beta = g$

(f) $\tau_0 = 3/4$, H_{1x}^N, $g_\alpha = g/5$, $g_\beta = g$

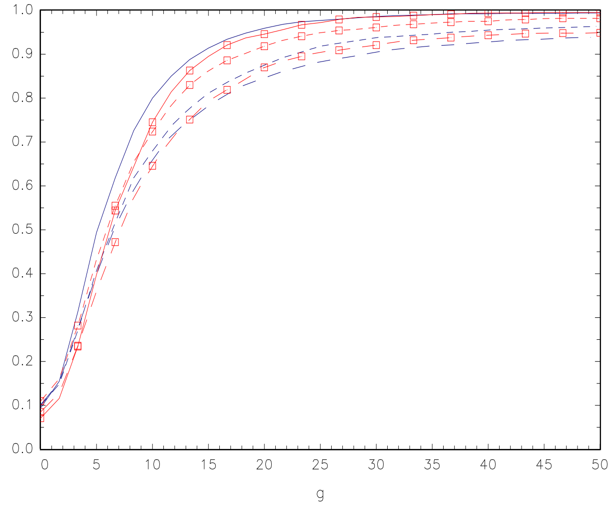
Figure 2. Asymptotic local power of tests under H^N: LM_1^B: - - - , LM_x^B: — — , LM_{1x}^B: —— , $SupF_1^B$: -□-, $SupF_x^B$: —□— , $SupF_{1x}^B$: —□—



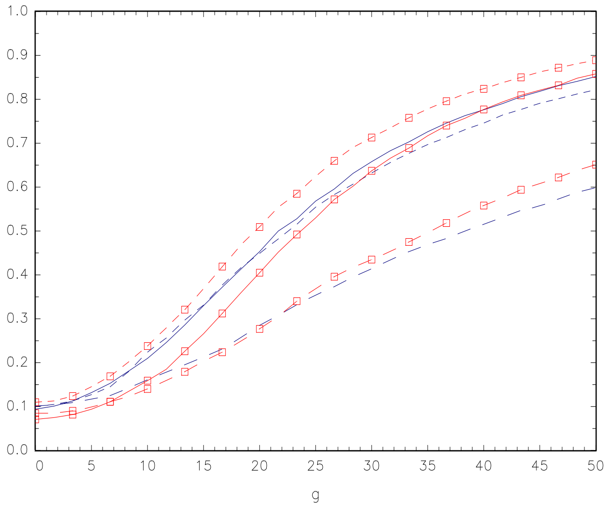
(a) $c_\alpha = c_\beta = 0$, H_1^S, $g_\alpha = 3g/5$



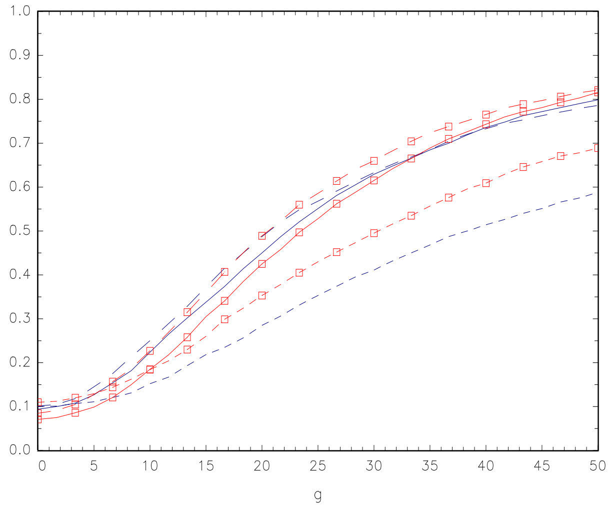
(b) $c_\alpha = c_\beta = 0$, H_x^S, $g_\beta = 3g$



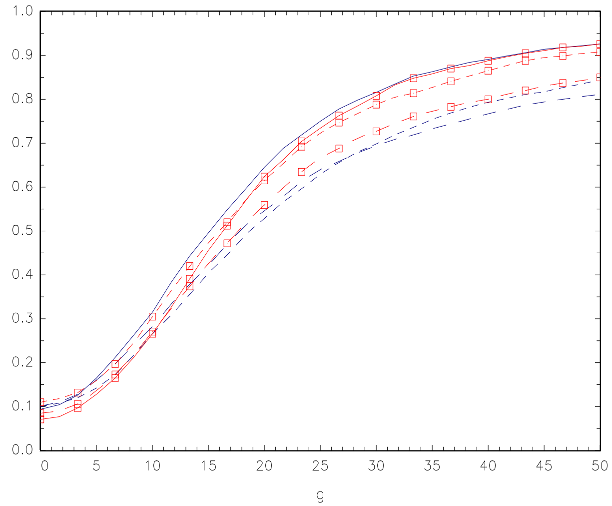
(c) $c_\alpha = c_\beta = 0$, H_{1x}^S, $g_\alpha = 3g/5$, $g_\beta = 3g$



(d) $c_\alpha = c_\beta = 10$, H_1^S, $g_\alpha = 3g/5$

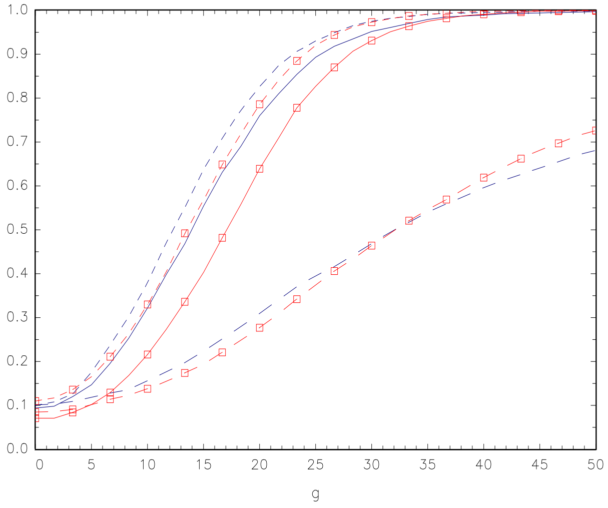


(e) $c_\alpha = c_\beta = 10$, H_x^S, $g_\beta = 3g$

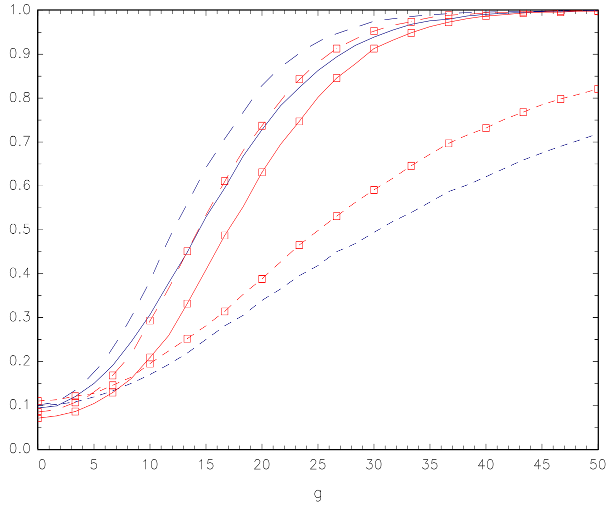


(f) $c_\alpha = c_\beta = 10$, H_{1x}^S, $g_\alpha = 3g/5$, $g_\beta = 3g$

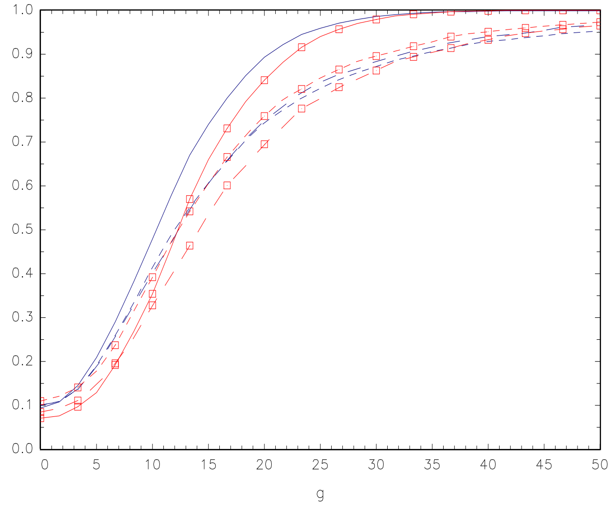
Figure 3. Finite sample power of tests under H^S: LM_1^B: - - - , LM_x^B: — — , LM_{1x}^B: —— , $SupF_1^B$: -□-, $SupF_x^B$: —□— , $SupF_{1x}^B$: —□—



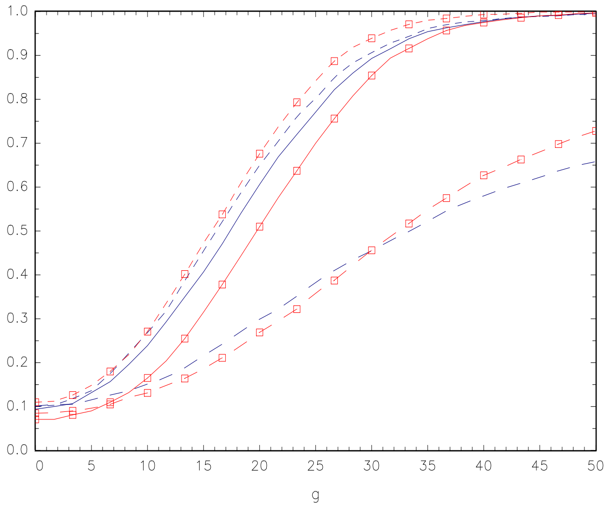
(a) $\tau_0 = 1/2$, H_1^N, $g_\alpha = g/5$



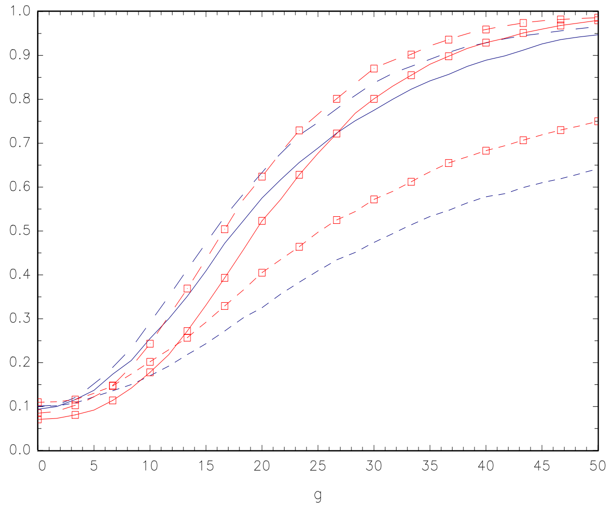
(b) $\tau_0 = 1/2$, H_x^N, $g_\beta = g$



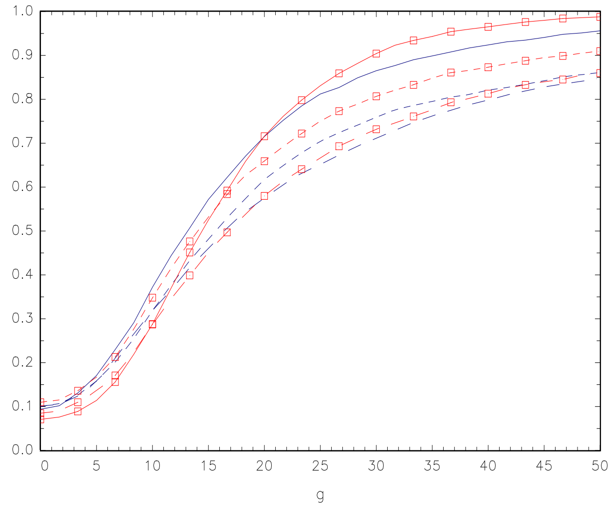
(c) $\tau_0 = 1/2$, H_{1x}^N, $g_\alpha = g/5$, $g_\beta = g$



(d) $\tau_0 = 3/4$, H_1^N, $g_\alpha = g/5$

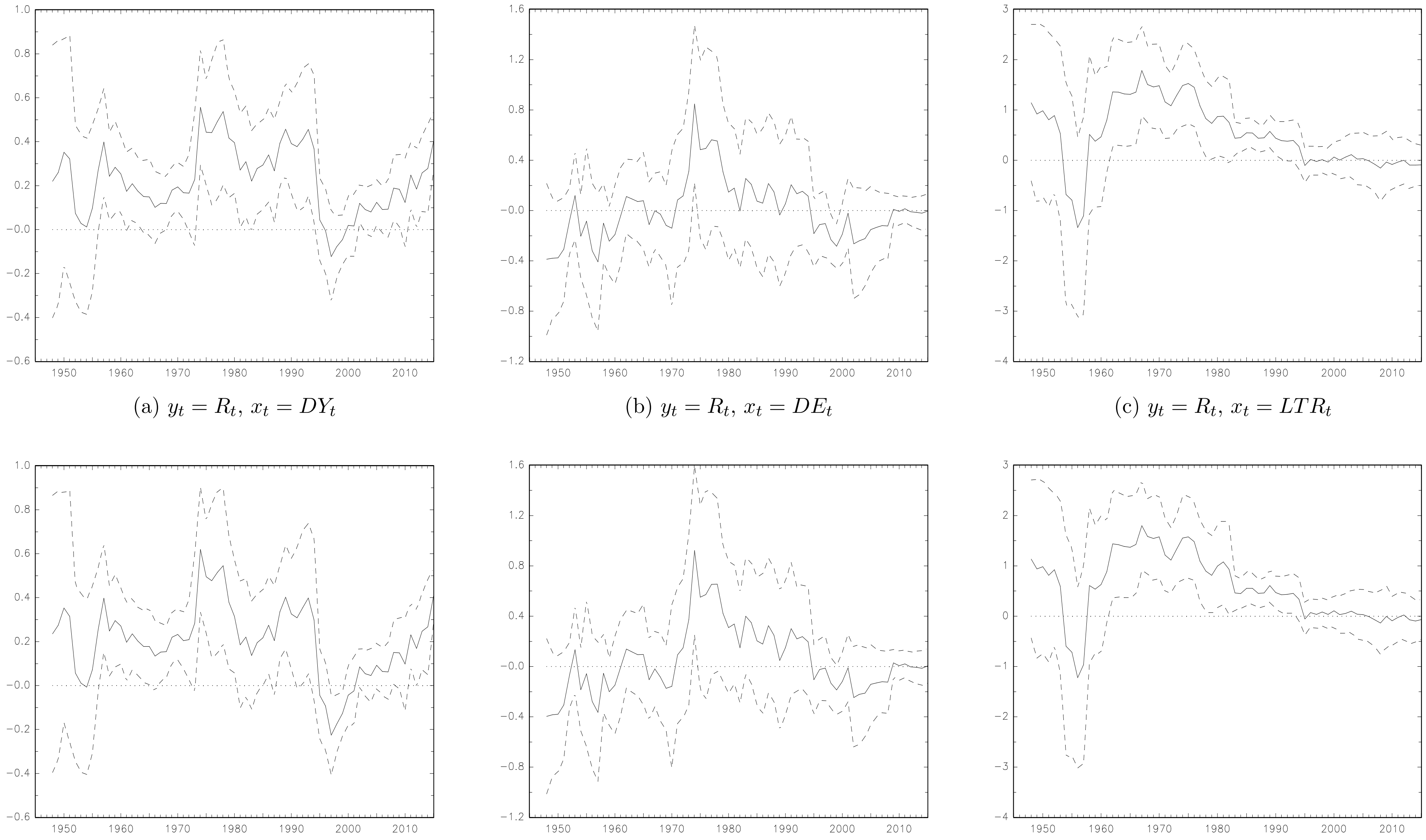


(e) $\tau_0 = 3/4$, H_x^N, $g_\beta = g$



(f) $\tau_0 = 3/4$, H_{1x}^N, $g_\alpha = g/5$, $g_\beta = g$

Figure 4. Finite sample power of tests under H^N: LM_1^B: - - - , LM_x^B: — — , LM_{1x}^B: —— , $SupF_1^B$: -⊡-, $SupF_x^B$: —⊡—, $SupF_{1x}^B$: —⊡—



(a) $y_t = R_t$, $x_t = DY_t$ (b) $y_t = R_t$, $x_t = DE_t$ (c) $y_t = R_t$, $x_t = LTR_t$

(d) $y_t = EP_t$, $x_t = DY_t$ (e) $y_t = EP_t$, $x_t = DE_t$ (f) $y_t = EP_t$, $x_t = LTR_t$

Figure 5. Rolling predictive regression IV coefficient estimates and standard error bands:
$y_t = \hat{\alpha} + \hat{\beta} x_{t-1}$; $\hat{\beta}$: —, $\pm 1.645 s.e.$: - - -

Supplementary Online Appendix

to

"Testing for Parameter Instability in Predictive Regression Models"

by

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

Date: January 8, 2018

Contents: Section S.1 of this supplement contains alternative forms of the large sample results given in Theorems 1 and 2 of the paper in terms of scalar limiting processes. Section S.2 provides proofs for the results given in section S.1.

S.1 Scalar Limiting Representations

In Corollaries S.1 and S.2 we first provide explicit expressions in terms of scalar processes for the LM_{1x}, LM_1 and LM_x statistics, and for the $SupF_1$ and $SupF_x$ statistics, respectively. These representations allow us to see more clearly the dependence of these limiting distributions on the local-to-unity parameter driving x_t and, under local alternatives, on the form of local parameter instability present.

Before stating Corollary S.1 we first need some additional definitions. Define the demeaned analogues of $B_{\eta 1,c_x}(r)$ and $M_{\eta\alpha,c_\alpha}(r)$, viz., $\bar{B}_{\eta 1,c_x}(\cdot) := B_{\eta 1,c_x}(\cdot) - \int_0^1 B_{\eta 1,c_x}(s)ds$ and $\bar{M}_{\eta\alpha,c_\alpha}(\cdot) := M_{\eta\alpha,c_\alpha}(\cdot) - \int_0^1 M_{\eta\alpha,c_\alpha}(s)ds$, respectively, and additionally define $\bar{M}_{\eta\beta,\eta x,c_\beta,c_x}(\cdot) := M_{\eta\beta,c_\beta}(\cdot)M_{\eta x,c_x}(\cdot) - \int_0^1 M_{\eta\beta,c_\beta}(s)M_{\eta x,c_x}(s)ds$.

Corollary S.1. *Let the conditions of Theorem 1 hold. Then:*

(i) ***S:*** *When $s_{\alpha t}$ and $s_{\beta t}$ are generated according to (4), with $a = g_\alpha T^{-1}$ and $b = g_\beta T^{-3/2}$,*

$$LM_1 \xrightarrow{w} \mathcal{LM}_1^S, \;\; LM_x \xrightarrow{w} \mathcal{LM}_x^S, \;\; \text{and} \;\; LM_{1x} \xrightarrow{w} \mathcal{LM}_1^S + \mathcal{LM}_{\bar{x}}^S$$

where

$$\mathcal{LM}_1^S := \int_0^1 \left[H_0(r) + \left\{ \int_0^1 d_2^2(s) \right\}^{-1/2} H_g^S(r) \right]^2$$

$$\mathcal{LM}_x^S := \frac{1}{\int_0^1 B_{\eta 1,c_x}(r)^2} \int_0^1 \left[\int_0^r B_{\eta 1,c_x}(s)dH_0(s) + \left\{ \int_0^1 d_2^2(s) \right\}^{-1/2} \int_0^r B_{\eta 1,c_x}(s)dH_g^S(s) \right]^2$$

$$\mathcal{LM}_{\bar{x}}^S := \frac{1}{\int_0^1 \bar{B}_{\eta 1,c_x}(r)^2} \int_0^1 \left[\int_0^r \bar{B}_{\eta 1,c_x}(s)dH_0(s) + \left\{ \int_0^1 d_2^2(s) \right\}^{-1/2} \int_0^r \bar{B}_{\eta 1,c_x}(s)dH_g^S(s) \right]^2$$

with

$$H_0(s) := \mathbb{B}_{\eta 2}(s) - \frac{\int_0^s \bar{B}_{\eta 1,c_x}(r)}{\int_0^1 \bar{B}_{\eta 1,c_x}(r)^2} \int_0^1 \bar{B}_{\eta 1,c_x}(r) dB_{\eta 2}(r)$$

where $\mathbb{B}_{\eta 2}(s) := B_{\eta 2}(s) - sB_{\eta 2}(1)$ *is the tied-down version of* $B_{\eta 2}(s)$*, and*

$$\begin{aligned} H_g^S(s) := \quad & g_\alpha \int_0^s \bar{M}_{\eta\alpha,c_\alpha}(r) + g_\beta \int_0^s \bar{M}_{\eta\beta,\eta x,c_\beta,c_x}(r) \\ & - \frac{\int_0^s \bar{B}_{\eta 1,c_x}(r)}{\int_0^1 \bar{B}_{\eta 1,c_x}(r)^2} \{ g_\alpha \int_0^1 \bar{B}_{\eta 1,c_x}(r) M_{\eta\alpha,c_\alpha}(r) + g_\beta \int_0^1 \bar{B}_{\eta 1,c_x}(r) M_{\eta\beta,\eta x,c_\beta,c_x}(r) \}. \end{aligned}$$

(ii) ***N:*** *When* $s_{\alpha t}$ *and* $s_{\beta t}$ *are generated according to (5), with* $a = g_\alpha T^{-1/2}$ *and* $b = g_\beta T^{-1}$*,*

$$LM_1 \xrightarrow{w} \mathcal{LM}_1^N, \;\; LM_x \xrightarrow{w} \mathcal{LM}_x^N, \;\; \text{and} \;\; LM_{1x} \xrightarrow{w} \mathcal{LM}_1^N + \mathcal{LM}_{\bar{x}}^N$$

where

$$\begin{aligned} \mathcal{LM}_1^N &:= \int_0^1 \left[H_0(r) + \left\{ \int_0^1 d_2^2(s) \right\}^{-1/2} H_g^N(r) \right]^2 \\ \mathcal{LM}_x^N &:= \frac{1}{\int_0^1 B_{\eta 1,c_x}(s)^2} \int_0^1 \left[\int_0^r B_{\eta 1,c_x}(s) dH_0(s) + \left\{ \int_0^1 d_2^2(s) \right\}^{-1/2} \int_0^r B_{\eta 1,c_x}(s) dH_g^N(s) \right]^2 \\ \mathcal{LM}_{\bar{x}}^N &:= \frac{1}{\int_0^1 \bar{B}_{\eta 1,c_x}(s)^2} \int_0^1 \left[\int_0^r \bar{B}_{\eta 1,c_x}(s) dH_0(s) + \left\{ \int_0^1 d_2^2(s) \right\}^{-1/2} \int_0^r \bar{B}_{\eta 1,c_x}(s) dH_g^N(s) \right]^2 \end{aligned}$$

and

$$\begin{aligned} H_g^N(s) \;\; := \;\; & 1(s \geq \tau_0) \left\{ g_\alpha(s - \tau_0) + g_\beta \int_{\tau_0}^s M_{\eta x,c_x}(r) \right\} - s \left\{ g_\alpha(1 - \tau_0) + g_\beta \int_{\tau_0}^1 M_{\eta x,c_x}(r) \right\} \\ & - \frac{\int_0^s \bar{B}_{\eta 1,c_x}(r)}{\int_0^1 \bar{B}_{\eta 1,c_x}(r)^2} \int_{\tau_0}^1 \bar{B}_{\eta 1,c_x}(r) \{ g_\alpha + g_\beta M_{\eta x,c_x}(r) \} \end{aligned}$$

Corollary S.2. *Let the conditions of Theorem 1 hold. Then, if either:*

(i) ***S:*** $s_{\alpha t}$ *and* $s_{\beta t}$ *are generated according to (4), with* $a = g_\alpha T^{-1}$ *and* $b = g_\beta T^{-3/2}$*, or*

(ii) ***N:*** $s_{\alpha t}$ *and* $s_{\beta t}$ *are generated according to (5), with* $a = g_\alpha T^{-1/2}$ *and* $b = g_\beta T^{-1}$*, then*

$$\begin{aligned} SupF_1 &\xrightarrow{w} \sup_{r \in \Lambda} \frac{[H_0(r) + \{\int_0^1 d_2(s)^2\}^{-1/2} H_g^A(r)]^2}{r - r^2 - \{\int_0^1 \bar{B}_{\eta 1,c_x}^2(s)\}^{-1} \{\int_0^r \bar{B}_{\eta 1,c_x}(s)\}^2}, \\ SupF_x &\xrightarrow{w} \sup_{r \in \Lambda} \frac{[\int_0^r B_{\eta 1,c_x}(s) dH_0(s) + \{\int_0^1 d_2(u)^2\}^{-1/2} \int_0^r B_{\eta 1,c_x}(s) dH_g^A(s)]^2}{\int_0^r B_{\eta 1,c_x}^2(s) - \{\int_0^r B_{\eta 1,c_x}(s)\}^2 - \{\int_0^1 \bar{B}_{\eta 1,c_x}^2(s)\}^{-1} \{\int_0^r \bar{B}_{\eta 1,c_x}(s) B_{\eta 1,c_x}(s)\}^2} \end{aligned}$$

where the superscript $A \in \{N, S\}$ *indicates either scheme* **N** *or scheme* **S***, and all other notation is as defined in Corollary S.1.*

Remark S.1. The limit expressions given in Corollaries S.1 and S.2 clearly show how g_α and g_β both enter the asymptotic distributions of all of the structural change statistics under each of schemes **S** and **N**. It is the presence of these terms that is seen to be the source of power for the tests based on these statistics to distinguish between H_0 and the various alternative hypotheses. □

Remark S.2. Under H_0, where $g_\alpha = g_\beta = 0$, the limiting distribution of LM_{1x}, under both schemes **S** and **N** can be written in the form

$$\int_0^1 H_0(r)^2 + \frac{1}{\int_0^1 \bar{B}_{\eta 1,c_x}(r)^2} \int_0^1 \left\{ \int_0^r \bar{B}_{\eta 1,c_x}(s) dH_0(s) \right\}^2$$

from which it can be seen that the limiting null distribution of LM_{1x} depends on any unconditional heteroskedasticity present in ϵ_{xt} and ϵ_{yt} and on the local-to-unity parameter, c_x, but does not depend on h_{21}, and, hence, does not depend on the correlation between ϵ_{xt} and ϵ_{yt}. The same comments can also be seen to apply to the limiting null distributions of the LM_1 and LM_x statistics and to those of the $SupF_{1x}$, $SupF_1$ and $SupF_x$ statistics. □

Remark S.3. Regarding the LM_1 statistic, a comparison of the result in Corollary S.1 with Theorem 1 of Shin (1994,pp.95-96) shows that the limiting null distribution of LM_1 coincides with the limiting null distribution of the CI_μ statistic for testing the null of co-integration in Shin (1994,p.95) for the case of a single pure $I(1)$ regressor (so that $c_x = 0$) and where ϵ_{xt} and ϵ_{yt} are homoskedastic. The result for LM_1 in Corollary 1 therefore also provides the limiting null representation for Shin's CI_μ statistic in the case where ϵ_{xt} and ϵ_{yt} satisfy Assumption 1, and indeed in the case where x_t is near-integrated. In the case considered in section 6.2 this coincidence would hold with the limiting null distribution given for the CI_τ statistic in Theorem 1 of Shin (1994) for the case of k regressors in the case where $\mathbf{f}_t := (1, t)'$ and would generalise that result to a general deterministic case more generally, again additionally allowing for near unit root behaviour in the regressors and heteroskedasticity in the innovations of the form considered in Assumption 1. It should be stressed again, however, that LM_1 and Shin's CI statistics are implemented in distinct contexts; see again the discussion at the beginning of section 2 on this point. □

Remark S.4. The limiting distribution given in Corollary 1 for LM_1 under scheme **S** coincides with the limiting distribution given in Theorem 3 of GHLT for the particular choice of $s_{\alpha t}$ made in GHLT. □

We next provide scalar versions of the large sample results given in Theorem 2 of the paper. In particular, we provide the conditional counterpart of the results give in Corollary S.1 above. A conditional counterpart of Corollary S.2 can be formulated in an entirely analogous fashion and is therefore omitted in the interests of brevity.

Corollary S.3. *Let the conditions of Theorem 2 hold. Then the following converge jointly as $T \to \infty$, in the sense of joint weak convergence of random measures on $\mathbb{R}$:*

$$LM_{1x}|\, x \xrightarrow{w} \int_0^1 \left[H_0(r) + \left\{ \int_0^1 d_2^2(s) \right\}^{-1/2} H_g^A(r) \right]^2$$
$$+ \frac{1}{\int_0^1 \bar{B}_{\eta 1, c_x}(r)^2} \int_0^1 \left[\int_0^r \bar{B}_{\eta 1, c_x}(s) dH_0(s) + \left\{ \int_0^1 d_2^2(s) \right\}^{-1/2} \int_0^r \bar{B}_{\eta 1, c_x}(s) dH_g^A(s) \right]^2 \Bigg| B_1$$

where the superscript $A \in \{N, S\}$ indicates either scheme **N** *or scheme* **S***, and*

$$LM_{1x}^* \mid x, y \xrightarrow{w} \int_0^1 \{H_0(r)\}^2 + \frac{1}{\int_0^1 \bar{B}_{\eta 1, c_x}(s)^2} \int_0^1 \left\{ \int_0^r \bar{B}_{\eta 1, c_x}(s) dH_0(s) \right\}^2 \Bigg| B_1.$$

Remark S.5. For the bootstrap LM_{1x}^*, LM_1^* and LM_x^* statistics the stated limiting distributions can all be seen not to depend on g_α and g_β; that is, the same limiting distributions are obtained for each of these statistics under the local alternatives under consideration as are obtained under the null hypothesis. In contrast, for LM_{1x}, LM_1 and LM_x (conditional on x), a stochastic offset, arising from the terms involving g_α and g_β, is seen in the limiting distributions. It is important to note, however, that these limiting distributions are not the same as those which appear for these statistics in Theorem 1. The implication of this is that while the bootstrap tests will have non-trivial asymptotic local power functions, these will not coincide (for a given alternative) with the asymptotic local power functions of (infeasible) versions of the LM_{1x}, LM_1 and LM_x tests based on knowledge of the unknown parameter, c_x. The same comments apply to the analogous *SupF*-based procedures. □

Remark S.6. Under conditional homoskedasticity of ϵ_{yt} corrected for ϵ_{xt} (i.e., under $d_{2t} = 1$), the process $H_0(\cdot)$ is a generalised Brownian bridge conditionally on $B_1(\cdot)$. Specifically, it is distributed like $B_2(\cdot)$ conditioned on $B_1(\cdot)$, $B_2(1) = 0$ and $\int_0^1 B_{\eta 1, c_x}(r) dB_2(r) = 0$ (Sottinen and Yazigi (2014), Theorem 3.1). As a result, the common random limit distribution of LM_{1x}^* given the data and, under H_0, of LM_{1x}, can be represented as the (regular) conditional distribution

$$\int_0^1 B_2^2(r) + \frac{1}{\int_0^1 \bar{B}_{\eta 1, c_x}(r)^2} \int_0^1 \left\{ \int_0^r \bar{B}_{\eta 1, c_x} dB_2 \right\}^2 \Bigg| \left\{ B_1, B_2(1) = 0, \int_0^1 B_{\eta 1, c_x} dB_2 = 0 \right\}.$$

The conditions $B_2(1) = 0$ and $\int_0^1 B_{\eta 1, c_x}(r) dB_2(r) = 0$ have a natural interpretation as limits of corresponding finite-sample orthogonality conditions satisfied by OLS residuals and explain the appearance of the process $H_0(.)$ in the limiting expressions. Similar considerations apply to the remaining statistics under discussion. The conditional representation of the limit distributions extends to the alternative hypothesis H_1 by viewing also $H_g^N(\cdot)$ and $H_g^S(\cdot)$ as representations of conditioned processes. □

S.2 Proofs of the Results in Section S.1

Proof of Corollary S.1

Because $\mathbf{a}_t$ contains a constant term, it can easily be seen that LM_{1x} is invariant to transformations of the form $\mathbf{a}_t \to \mathbf{a}_t - [0\ \xi]' = [1\ (x_{t-1} - \xi)]'$ for any r.v. ξ. Then, choosing $\xi = \bar{x}_{-1}$ where $\bar{x}_{-1} := T^{-1}\sum_{t=1}^{T} x_{t-1}$, we have that

$$\begin{aligned} LM_{1x} &= \frac{1}{T\hat{\sigma}^2}\sum_{i=1}^{T}\left(\sum_{t=1}^{i}[1\ \mathring{x}_{t-1}]\hat{e}_t\right)\left(\sum_{t=1}^{T}[1\ \mathring{x}'_{t-1}][1\ \mathring{x}_{t-1}]\right)^{-1}\left(\sum_{t=1}^{i}[1\ \mathring{x}_{t-1}]'\hat{e}_t\right) \\ &= LM_1 + LM_{\bar{x}}, \end{aligned}$$

where LM_1 is as defined in (8) while $LM_{\bar{x}} := \frac{1}{T\hat{\sigma}^2\sum_{t=1}^{T}\mathring{x}_{t-1}^2}\sum_{i=1}^{T}\left(\sum_{t=1}^{i}\mathring{x}_{t-1}\hat{e}_t\right)^2$. The expression for the weak limit of the LM_{1x} statistics decomposes conformably with this finite-sample decomposition; that is,

$$\begin{aligned} LM_{1x} &\overset{w}{\to} \frac{1}{\int_0^1 d_2^2(r)}\int_0^1 H(r)^2 + \frac{1}{\int_0^1 d_2^2(r)\int_0^1 \bar{M}_{\eta x,c_x}(r)^2}\int_0^1\{\int_0^r \bar{M}_{\eta x,c_x}(s)dH(s)\}^2 \\ &= \frac{1}{\int_0^1 d_2^2(r)}\int_0^1 H(r)^2 + \frac{1}{\int_0^1 d_2^2(r)\int_0^1 \bar{B}_{\eta 1,c_x}(r)^2}\int_0^1\{\int_0^r \bar{B}_{\eta 1,c_x}(s)dH(s)\}^2. \end{aligned}$$

By further decomposing $H(r) = \{\int_0^1 d_2(r)^2\}^{1/2}H_0(r) + H_g^A(r)$, where the superscript $A \in \{N, S\}$ indicates either scheme **N** or scheme **S**, the results in Corollary S.1 are obtained. The limits of LM_1 and LM_x follow by considering obvious linear transformations of $\mathring{\mathbf{a}}_{t-1}$ from the proof of Theorem 1. ■

Proof of Corollary S.2

The result follows from the following general observation. Consider the *SupF* statistic, say $SupF_{(u,v)}$, for testing the null hypothesis $H_0 : \varphi = 0$ in the regression

$$y_t = \alpha + \beta x_{t-1} + \varphi D_t(\lfloor \tau T \rfloor)(u + v x_{t-1}) + \beta_0 \Delta x_t + error_t, \qquad t = 1, ..., T,$$

where u, v are known and not simultaneously equal to zero. As $u + v x_{t-1} = (u, v)\,\mathbf{a}_{t-1}$, by relating $\mathbf{a}_{t-1}$ to $\mathring{\mathbf{a}}_{t-1}$ it follows that, under the conditions of Theorem 1, the limit distribution of $SupF_{(u,v)}$ can be represented as

$$SupF_{(u,v)} \overset{w}{\to} \sup_{r\in\Lambda} \frac{(u,v)\,\boldsymbol{\mu}\mathbf{J}(r)\mathbf{J}'(r)\boldsymbol{\mu}'\,(u,v)'}{(u,v)\,\boldsymbol{\mu}\{\mathbf{V}(r) - \mathbf{V}(r)\,\mathbf{V}(1)^{-1}\mathbf{V}(r)\}\boldsymbol{\mu}'\,(u,v)'}, \tag{S.1}$$

where $\boldsymbol{\mu} = \{\boldsymbol{\mu}_{ij}\}_{i,j=1}^2$ has $\boldsymbol{\mu}_{11} = \boldsymbol{\mu}_{22} = 1$, $\boldsymbol{\mu}_{12} = 0$ and $\boldsymbol{\mu}_{21} = \int_0^1 M_{\eta x,c_x}(r)$. The stated limiting distributions for $SupF_1$ and $SupF_x$ are therefore obtained setting $(u, v) = (1, 0)$ and $(u, v) = (0, 1)$, respectively. ■

Proof of Corollary S.3

From the proof of Theorem 2, by defining $B_{\eta 2}^{\dagger} := (\int d_2^2)^{-1/2} B_m^*$, it follows that

$$LM_{1x}^* \mid x, y \xrightarrow{w} \int_0^1 \left\{H^{\dagger}(r)\right\}^2 + \frac{1}{\int_0^1 \bar{B}_{\eta 1, c_x}(s)^2} \int_0^1 \left\{\int_0^r \bar{B}_{\eta 1, c_x}(s) dH^{\dagger}(s)\right\}^2 \Bigg| B_1$$

where

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

It remains to observe that $(H_0, B_1) \stackrel{d}{=} \left(H^{\dagger}, B_1\right)$. ■

Reference

Sottinen T. and A. Yazigi (2014). Generalized Gaussian bridges. *Stochastic Processes and Their Applications* 124, 3084-3105.