

Improved tests for stock return predictability

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

To cite this article: David I. Harvey, Stephen J. Leybourne & A. M. Robert Taylor (2023): Improved tests for stock return predictability, *Econometric Reviews*, DOI: [10.1080/07474938.2023.2222634](https://doi.org/10.1080/07474938.2023.2222634)

To link to this article: <https://doi.org/10.1080/07474938.2023.2222634>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC



Published online: 14 Jul 2023.



Submit your article to this journal [↗](#)



Article views: 56



View related articles [↗](#)



View Crossmark data [↗](#)

Improved tests for stock return predictability

David I. Harvey^a, Stephen J. Leybourne^a, and A. M. Robert Taylor^b

^aSchool of Economics, University of Nottingham, Nottingham, UK; ^bEssex Business School, University of Essex, Colchester, UK

ABSTRACT

Predictive regression methods are widely used to examine the predictability of (excess) stock returns by lagged financial variables characterized by unknown degrees of persistence and endogeneity. We develop a new hybrid test for predictability in these circumstances based on simple regression t -statistics. Where the predictor is endogenous, the optimal, but infeasible, test for predictability is based on the t -statistic on the lagged predictor in the basic predictive regression augmented with the current period innovation driving the predictor. We propose a feasible version of this augmented test, designed for the case where the predictor is an endogenous near-unit root process, using a GLS-based estimate of the innovation used in the infeasible test regression. The limiting null distribution of this statistic depends on both the endogeneity correlation parameter and the local-to-unity parameter characterizing the predictor. A method for obtaining asymptotic critical values is discussed and response surfaces are provided. We compare the asymptotic power properties of the feasible augmented test with those of a (non augmented) t -test recently considered in Harvey et al. and show that the augmented test is more powerful in the strongly persistent predictor case. We then propose using a weighted combination of the augmented statistic and the t -statistic of Harvey et al., where the weights are obtained using the p -values from a unit root test on the predictor. We find this can further improve asymptotic power in cases where the predictor has persistence at or close to that of a unit root process. Our final hybrid testing procedure then embeds the weighted statistic within a switching-based procedure which makes use of a standard predictive regression t -test, compared with standard normal critical values, when there is evidence for the predictor being weakly persistent. Monte Carlo simulations suggest that overall our new hybrid test displays superior finite sample performance to comparable extant tests.

ARTICLE HISTORY

Received 11 January 2022
Accepted 06 March 2023

KEYWORDS

Augmented regression;
endogeneity; persistence;
predictive regression;
weighted statistics

JEL CLASSIFICATION

C12, C22

1. Introduction

Many studies in the applied economics and finance literature have focused on testing for the predictability of asset returns, employing a range of candidate predictor variables, such as valuation ratios, interest rates, and other financial and macroeconomic variables. By way of examples, Fama (1981) considers various predictors including interest rates, industrial production, GNP, and capital stock and expenditure, while Campbell and Yogo (2006) consider the dividend-price ratio, the earnings-price ratio, the three-month T-bill rate, and the long-short yield spread. Standard approaches to testing predictability are based on a simple linear regression model with a constant and lagged putative predictor (x_{t-1} say), with a corresponding regression coefficient β .

CONTACT A. M. Robert Taylor  robert.taylor@essex.ac.uk  Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK.

© 2023 The Author(s). Published with license by Taylor and Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

In empirical studies, it is commonly found that the candidate predictor variables are highly persistent (with either unit root or near unit root autoregressive processes) and also endogenous with a non zero (often strongly negative) correlation between the errors in the predictive regression and the innovations driving the predictor process; see, *inter alia*, Campbell and Yogo (2006), Goyal and Welch (2003), and Welch and Goyal (2008). In the presence of strong persistence and endogeneity, Cavanagh et al. (1995) show that the standard t -test on the estimate of β suffers from severe size distortions; see also Campbell and Yogo (2006), Nelson and Kim (1993), and Stambaugh (1999). This finding has motivated the development of tests for predictability that are designed to allow for both endogeneity and strong persistence in the predictor series x_t , modeled by a first-order autoregression with a local-to-unity coefficient $\phi = 1 - cT^{-1}$ (where c is an unknown finite constant and T is the sample size).

As a result, a number of likelihood-based predictability tests have been developed in the literature which are designed to be asymptotically valid when the predictor is strongly persistent and endogenous; see, *inter alia*, Cavanagh et al. (1995), Lewellen (2004), Campbell and Yogo (2006), Elliott et al. (2015) [EMW, hereafter], and Jansson and Moreira (2006), and most recently a hybrid test, based around a number of simple regression t -ratios, developed in Harvey et al. (2021) [HLT hereafter]. Arguably the most widely applied of these tests in the literature is the Q test of Campbell and Yogo (2006), which falls within the general control variable approach outlined in Elliott (2011). Here the simple linear predictive model is augmented by an additional regressor used as a proxy for the current period innovation driving the predictor; an infeasible version of this test using the actual current period innovation is optimal when the predictor is endogenous. In particular, in its simplest form, Q is based around the infeasible t -statistic on β when $(x_t - \phi x_{t-1})$ is added as a regressor to the predictive regression. Campbell and Yogo (2006) develop a feasible version of this test, using the approach of Cavanagh et al. (1995), based on a Bonferroni confidence interval for β obtained using a confidence interval for ϕ (equivalently c) formed from the well-known quasi-GLS demeaned augmented Dickey-Fuller [ADF] unit root statistic of Elliott et al. (1996).

Among the likelihood-based approaches listed above, only the procedures developed in EMW and HLT are also asymptotically valid for the case of a weakly persistent predictor.¹ Like Lewellen (2004), the testing procedure outlined in EMW rules out the possibility that the predictor x_t is locally explosive (by imposing that c is non negative), while HLT and Campbell and Yogo (2006) allow for some local explosivity ($-5 \leq c < 0$) in the predictor. Simulation results presented in HLT suggest that where the predictor is locally explosive the Q test of Campbell and Yogo (2006), although valid, displays very poor power and is easily dominated by the hybrid test proposed in HLT; while the EMW test is highly unreliable. Where the possibility of local explosivity in the predictor can be ruled out, based on their simulation results HLT find that the EMW test dominates other tests where the predictor is either a pure unit root process ($c = 0$) or lies very close to a unit root process (c is small and positive) arguing that "... it appears that exclusion of robustness to the case of explosive predictors affords the EMW test the opportunity of greater power in the unit root setting..." *op. cit.* p.207. For larger c , HLT argue on the basis of their simulations that their proposed hybrid test offers superior power to all of the leading tests in the literature, including the EMW test.

Our aim in this article is to investigate an alternative to the hybrid testing procedure of HLT designed to exploit available power advantages that exist for strongly persistent predictors when c is either zero or small and positive in cases where locally explosive predictors can be ruled out, *a priori*. This then allows us to develop a procedure that can be compared on a level playing field with the EMW test. The approach we outline will be focused on easy to implement tests based on regression t -ratios. The hybrid testing procedure we propose can be viewed as an extension of the hybrid test outlined in HLT with the

¹A different strand of the literature which allows for both weakly and strongly persistent predictors is characterized by contributions from Phillips and Magdalinos (2009), Kostakis et al. (2015) and Breitung and Demetrescu (2015) and focuses on instrumental variable [IV] estimation using an instrument constructed from the predictor variable and designed to be less persistent than a local-to-unity process. While such IV based tests are valid regardless of the degree of persistence in the predictor, they are less powerful than the tests of EMW and HLT, particularly so when the predictor is weakly persistent or where it is strongly persistent with c zero or close to zero; see Figures 2–6 in HLT pp.208-212.

introduction of information from an additional t -ratio motivated by the control variable approach of Elliott (2011). This t -ratio is formed on the lagged predictor in the basic predictive regression augmented with a GLS-based estimate used to proxy the current period innovation driving the predictor. In the strongly persistent case, we show that the limiting null distribution of this statistic depends on both the endogeneity correlation parameter and the local-to-unity parameter characterizing the predictor. We therefore propose a feasible method for obtaining asymptotically conservative critical values and provide response surfaces for practical use. An analysis of the asymptotic local power function of the resulting conservative test together with a corresponding feasible (conservative) implementation of the t -ratio proposed in HLT, obtained from a variant of the standard predictive regression where the OLS demeaned returns are regressed on the GLS demeaned lagged predictor, shows that, in the empirically most relevant case where a significant negative correlation exists between returns and the predictor's innovations, the new proxy-based test is more powerful than the corresponding test from HLT for positive predictability ($\beta > 0$) for $c = 0$ and small values of c ; that is, exactly the areas where the original hybrid test of HLT is less powerful than the EMW test. We then show that substantial further power improvements can be obtained in these scenarios by considering a weighted combination of the new t -ratio and the t -ratio from HLT, the weights depending on the persistence of the predictor via a function of the p -values from a standard Dickey-Fuller-type unit root test applied to the predictor, again made operational using asymptotically conservative critical values with a response surface provided for practical implementation.

Like HLT we find that when testing for positive persistence with a positive or small negative endogeneity correlation, or when testing for negative predictability ($\beta < 0$) with an endogeneity correlation that is not significantly positive, asymptotic local power is improved by using the standard predictive regression t -statistic with an asymptotically conservative critical value. Consequently, when testing for positive (negative) predictability, our recommended procedure in the near-unit root environment is to use the conservative standard t -ratio when the estimated endogeneity correlation is either positive or “small” and negative (either negative or “small” and positive), but to use the conservative test based on the weighted statistic otherwise. Further, in common with EMW and HLT, if the data suggest the predictor is weakly persistent, we propose switching into the standard t -ratio test with reference to standard normal critical values. Like HLT we base our switching function not on an (inconsistent) estimate of c , but rather on the familiar augmented Dickey-Fuller normalized bias coefficient unit root test.

In Monte Carlo simulations, we find that the hybrid test proposed in this article performs well in terms of finite sample size and power across a range of correlation parameters and persistence levels for the predictor, and compares very favorably with extant tests, offering a simple yet highly effective method for predictability testing. In particular, our proposed hybrid test almost always outperforms both the EMW and HLT hybrid test procedures in the case of strongly persistent predictors, with all three being largely identical for predictors displaying only very weak levels of persistence (as expected, given all three switch to a conventional t -test in this case). In cases where one is prepared to rule out the possibility of an explosive predictor, we therefore recommend the hybrid test developed in this article. Otherwise the hybrid test in HLT is preferred.

The remainder of the article is organized as follows. Section 2 introduces the predictive regression model which we will consider in this article together with the assumptions which we place on this data generating process [DGP]. In Section 3, we present the new augmented t -statistic that will subsequently feature in our hybrid testing procedure and detail its asymptotic properties. Here we also outline our method for obtaining asymptotic critical values and provide numerical comparisons with existing tests based on asymptotic local power functions. These simulation results provide motivation for the weighted statistics that we propose and evaluate in Section 4. Our final proposed hybrid testing procedure that allows for both weakly and strongly persistent predictors is then outlined in Section 5. Section 6 discusses extensions to deal with higher order serial correlation in the predictor. In Section 7, we investigate the finite sample size and power properties of our proposed hybrid test, comparing with the test procedures of EMW and HLT. Section 8 concludes. We use the notation $x := y$ ($x =: y$) to denote that x is defined by y (y is defined by x), and \Rightarrow to denote weak convergence.

2. The predictive regression model

Let y_t denote the (excess) stock return in period t and let x_{t-1} denote a variable observed at time $t - 1$ which is considered to be a putative predictor for y_t . The predictive regression model we consider is

$$y_t = \alpha_y + \beta x_{t-1} + \epsilon_{yt}, \quad t = 2, \dots, T \quad (1)$$

where x_t is an observed process, specified according to the DGP

$$\begin{aligned} x_t &= \alpha_x + s_t, & t &= 1, \dots, T \\ s_t &= \phi s_{t-1} + \epsilon_{xt}, & t &= 2, \dots, T \end{aligned} \quad (2)$$

with s_1 a mean zero $O_p(1)$ random variable.

As discussed in Section 1, it is important for practical purposes to allow for the possibility of high persistence in the predictor variable x_t and to allow the shocks driving the predictor, ϵ_{xt} in (2), to be correlated with the unpredictable component of stock returns, ϵ_{yt} in (1). As regards the latter, we assume that the innovation vector $\epsilon_t := (\epsilon_{xt}, \epsilon_{yt})'$ is IID with finite fourth-order moments and satisfying

$$\begin{bmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{bmatrix} \sim \text{IID} \left(0, \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} \right).$$

Remark 1. The assumption that ϵ_t is a vector IID process is made purely to simplify our presentation. All of the large sample results given in this article continue to hold in the case where ϵ_t is a (bivariate) martingale difference process satisfying the conditions given on p.200 of HLT. Indeed, for the case of a strongly persistent predictor (**Assumption S**), it is also possible to allow for conditional heteroskedasticity of the form considered in Assumption A.1 of Campbell and Yogo (2006) without altering the large sample results which are given in what follows. In the case of a weakly persistent predictor (**Assumption W**), the same would be true for conditional heteroskedasticity of the form given in, for example, Assumption INNOV(ii) of Kostakis et al. (2015, p.1512) providing the regression t -ratios discussed in what follows are implemented using White standard errors rather than OLS standard errors; notice, however, that for our final hybrid test outlined in Section 5 only the conventional t -test, \mathcal{J}_N , would actually need to be based on a t -statistic computed with White standard errors. The assumption that ϵ_{xt} is serially uncorrelated is also not crucial and we will subsequently discuss in Section 6 how the methods we propose can be modified to allow for weak dependence in ϵ_{xt} . The methods developed in the literature on predicting returns are, however, based on the assumption that ϵ_{yt} is serially uncorrelated; different methods are required in cases where ϵ_{yt} may be serially correlated and, as such, will not be considered here.

With respect to the degree of persistence in x_t , we assume that the true value of ϕ in (2) is unknown to the practitioner and satisfies one of the following two assumptions:

Assumption S. Strongly persistent predictor: The autoregressive parameter ϕ in (2) is local-to-unity with $\phi := 1 - cT^{-1}$ where c is a fixed non negative constant.

Assumption W. Weakly persistent predictor: The autoregressive parameter ϕ in (2) is fixed and bounded away from unity, $|\phi| < 1$.

Remark 2. Many putative predictors are strongly persistent, with sums of sample autoregressive coefficients close to or only slightly smaller than unity. In such cases, near-integrated asymptotics provide good approximations for the behavior of test statistics. However, not all possible predictors are strongly persistent and many models in the literature treat x_t as generated from a stable autoregressive process. We therefore allow for either of these possibilities to hold for x_t . As discussed in Section 1, our assumptions exclude the possibility of explosive predictors ($\phi > 1$), in line with the approach of, for example, EMW

and Lewellen (2004). In contrast, HLT and Campbell and Yogo (2006) both allow for a small degree of local explosivity ($-5 \leq c < 0$) in the predictor in the tests they develop.

In this article, our focus is on developing tests of the null hypothesis that y_t is not predictable by x_{t-1} , i.e., $H_0 : \beta = 0$ in (1), which do not require the practitioner to know which of [Assumption S](#) or [Assumption W](#) holds for ϕ in (2). The alternative hypothesis is that y_t is predictable by x_{t-1} , in which case $\beta > 0$ or $\beta < 0$ (one-sided alternatives are commonly adopted in practice). We will establish the large sample behavior of the predictability tests considered in this article under local alternatives such that the slope parameter β in (1) is local-to-zero. This approach permits analysis of the tests' local asymptotic power, and is consistent with the fact that predictive regressions for stock returns typically exhibit a small R^2 and low signal-to-noise ratios, with departures from the null being small when predictability is present. The appropriate localization rate (Pitman drift) is dictated by which of [Assumption S](#) and [Assumption W](#) holds. Under [Assumption S](#), where x_t is strongly persistent, the appropriate local alternative is given by $H_{1,S} : \beta = gT^{-1}$, while for weakly dependent x_t under [Assumption W](#), it is given by $H_{1,W} : \beta = gT^{-1/2}$, where in each case g is a finite constant.

The familiar Cholesky decomposition allows us to write the two components of ϵ_t in the form

$$\begin{aligned}\epsilon_{xt} &= \sigma_x e_{1t} \\ \epsilon_{yt} &= \sigma_y \left(\rho_{xy} e_{1t} + \sqrt{1 - \rho_{xy}^2} e_{2t} \right)\end{aligned}\quad (3)$$

where $e_t := (e_{1t}, e_{2t})' \sim IID(0, I_2)$ and $\rho_{xy} := \sigma_{xy}/(\sigma_x \sigma_y)$ is the contemporaneous correlation between the innovations driving the predictor, ϵ_{xt} , and the unpredictable component of stock returns, ϵ_{yt} . Using this representation, we can then re-write the predictive regression in (1) as

$$y_t = \alpha_y + \beta x_{t-1} + \left(\frac{\sigma_y}{\sigma_x} \rho_{xy} \right) \epsilon_{xt} + \left(\sigma_y \sqrt{1 - \rho_{xy}^2} \right) e_{2t}. \quad (4)$$

The representation in (4) is instructive, in that it demonstrates how a predictive regression featuring an endogenous predictor x_{t-1} , such as (1), can be re-written using ϵ_{xt} as an additional covariate in a form in which the predictor regressor, x_{t-1} , is strictly exogenous.

3. A new predictability test

In what follows it is convenient to define a generically notated regression model:

$$y_t = \alpha + \beta x_{t-1} + \delta z_{xt} + v_t \quad (5)$$

and consider the generic t -statistic associated with the OLS estimate of β in (5).

3.1. An infeasible test

If ϵ_{xt} was observed, which is equivalent to knowing ϕ (abstracting from the unknown constant, α_x), we could then perform a standard OLS regression in (5) with $z_{xt} = \epsilon_{xt}$, which is clearly a correct specification with respect to the DGP in (4). Denoting the corresponding infeasible t -statistic as \mathcal{T}^{inf} , it is straightforward to show that \mathcal{T}^{inf} has a standard normal limiting distribution under the null hypothesis H_0 , irrespective of whether [Assumption S](#) or [Assumption W](#) holds. Moreover, under Gaussianity this would be an efficient test (among α_y, α_x invariant tests) whenever $\rho_{xy} \neq 0$. Note that including ϵ_{xt} as a regressor reduces the error variance from σ_y^2 in (1) to $\sigma_y^2(1 - \rho_{xy}^2)$ in (4); that is, with knowledge of ϕ we can essentially subtract off the part of the innovation to returns that is correlated with the innovation to the predictor variable, thereby delivering a more powerful test. When $\rho_{xy} = 0$, \mathcal{T}^{inf} remains asymptotically efficient as incorporation of the redundant regressor ϵ_{xt} has no effect in large samples.

3.2. A feasible test using a proxy measure for ϵ_{xt}

Given that ϵ_{xt} is unobservable, one might ask if it is possible to obtain a proxy measure for ϵ_{xt} ? In fact, this testing problem falls within the general control variable approach outlined in Elliott (2011). Here, (1) is augmented by an additional regressor used as a proxy for the current period innovation driving the predictor, ϵ_{xt} ; so we may consider (5) as this augmented regression with z_{xt} the proxy regressor. There are a number of ways in which this approach can be implemented, including the Bonferroni-based method advocated in Campbell and Yogo (2006) which as discussed in Section 1 is based on a sequence of such augmented regressions. Here we consider an alternative approach based on a single regression including a covariate z_{xt} in (5) acting as a direct proxy for ϵ_{xt} in (4). We will also primarily focus our discussion on the case of Assumption S where x_t is strongly persistent, as this is the most problematic case where the standard t -statistic based on OLS estimation of (1), which we denote by \mathcal{T} , has a non pivotal limiting distribution. An obvious approach to obtaining a proxy for ϵ_{xt} is to assume a particular value for the local-to-unity parameter c , say \bar{c} ; we would then construct $z_{xt} = x_t - (1 - \bar{c}T^{-1})x_{t-1}$ (assuming $\alpha_x = 0$ for simplicity). If it happened to be the case that $\bar{c} = c$, then $z_{xt} = \epsilon_{xt}$ and we obtain the asymptotically standard normal and efficient test, \mathcal{T}^{inf} . However, when $\bar{c} \neq c$ the critical values for this test will depend on both ρ_{xy} and c , and it will no longer be an efficient test, with power being a (decreasing) function of the distance $|c - \bar{c}|$. This clearly poses a problem in implementation as c cannot be consistently estimated.

An obvious proxy for ϵ_{xt} is the OLS estimate, $\hat{\epsilon}_{xt}$ say, obtained from an OLS regression of Δx_t on a constant and x_{t-1} . However, setting $z_{xt} = \hat{\epsilon}_{xt}$ in (5) runs into the problem that z_{xt} is exact orthogonal to the predictive regressor x_{t-1} . The estimate of β from such a fitted model is then numerically identical to that which would be obtained if z_{xt} was omitted from (5). Moreover, the corresponding statistic is approximately $1/\sqrt{1 - \rho_{xy}^2}$ times the simple t -statistic, \mathcal{T} , and so the inference drawn from such a test would essentially be identical to that from \mathcal{T} , hence using the proxy regressor $\hat{\epsilon}_{xt}$ delivers no benefit whatsoever.

An alternative method for obtaining a proxy for ϵ_{xt} , which takes account of a strongly persistent autoregressive structure in estimating the intercept term α_x , is to employ a quasi-GLS estimate of α_x obtained from the quasi-differenced OLS regression of $(x_1, x_2 - \bar{\phi}x_1, \dots, x_T - \bar{\phi}x_{T-1})$ on $(1, 1 - \bar{\phi}, \dots, 1 - \bar{\phi})$ where $\bar{\phi} := 1 - \bar{c}/T$ with $\bar{c} = 7$; see Elliott et al. (1996) for further details. We denote this estimator $\tilde{\alpha}_x$. We would then estimate the OLS regression

$$\Delta x_t = \phi(x_{t-1} - \tilde{\alpha}_x) + \eta_t \quad (6)$$

and, denoting the estimate of ϕ by $\tilde{\phi}$, construct the residuals $\tilde{\epsilon}_{xt} := \Delta x_t - \tilde{\phi}(x_{t-1} - \tilde{\alpha}_x)$. Then we consider setting $z_{xt} = \tilde{\epsilon}_{xt}$ in (5). In contradistinction to the OLS-based proxy regressor $\hat{\epsilon}_{xt}$, the GLS-based proxy regressor $\tilde{\epsilon}_{xt}$ is not orthogonal to x_{t-1} . This lack of orthogonality raises the potential for $\tilde{\epsilon}_{xt}$ to act as a useful proxy for ϵ_{xt} in the strongly persistent case. We therefore construct the t -statistic associated with the OLS estimate of β in the regression

$$y_t = \alpha + \beta x_{t-1} + \delta \tilde{\epsilon}_{xt} + v_t \quad (7)$$

and denote this t -statistic as \mathcal{T}^* in what follows. As we shall establish in Section 3.3, the limiting null distribution of \mathcal{T}^* depends on both ρ_{xy} and c in the case where x_t is strongly persistent (Assumption S), although this issue notwithstanding, we might anticipate that this procedure could deliver decent power performance due to the inclusion of a proxy for ϵ_{xt} . Under Assumption W, the asymptotic distribution of $\tilde{\epsilon}_{xt}$, and therefore that of \mathcal{T}^* , will depend on the distribution of s_1 , so this statistic is appropriate only under Assumption S.

3.3. Asymptotic distribution of \mathcal{T}^*

In Theorem 1 we now report the asymptotic distribution of the \mathcal{T}^* statistic under both the null and local alternatives under $H_{1,S}$. A proof of Theorem 1 is provided in the appendix.

Theorem 1. Let y_t and x_t be generated according to the model in (1)-(2) under the conditions stated in Section 2 and let Assumption S hold. Then, as $T \rightarrow \infty$, under $H_{1,S}$:

$$\begin{aligned} \mathcal{T}_{xy}^* &\Rightarrow \frac{g\sigma_x}{\sigma_y} \frac{\sqrt{\int_0^1 \bar{W}_{1c}(r)^2 dr}}{\sqrt{1 - \rho_{xy}^2}} + \frac{\int_0^1 \bar{W}_{1c}(r) dW_2(r)}{\sqrt{\int_0^1 \bar{W}_{1c}(r)^2 dr}} + \frac{\rho_{xy} \sqrt{\int_0^1 \bar{W}_{1c}(r)^2 dr} \int_0^1 W_{1c}(r) dW_1(r)}{\sqrt{1 - \rho_{xy}^2} \int_0^1 W_{1c}(r)^2 dr} \\ &=: \mathcal{S}^*(g\sigma_x/\sigma_y, \rho_{xy}, c) \end{aligned}$$

where $W_1(r)$ and $W_2(r)$ are independent standard Brownian Motions, $\bar{W}_{1c}(r) := W_{1c}(r) - \int_0^1 W_{1c}(s) ds$ with $W_{1c}(r) := \int_0^r e^{-(r-s)c} dW_1(s)$.

Remark 3. The result in Theorem 1 highlights that the offset seen in the limiting distribution under the local alternative, $H_{1,S}$, given by the first term in the expression for $\mathcal{S}^*(g\sigma_x/\sigma_y, \rho_{xy}, c)$, is a function of the deterministic offset term $g\sigma_x/\sigma_y$, comprised of the Pitman drift, g , and the signal-to-noise ratio associated with the predictor, σ_x/σ_y , weighted by a stochastic offset term. Consequently, the test's asymptotic local power is higher, other things being equal, the larger the Pitman drift, and the larger the amount of variability in the predictor, relative to the error term in the predictive regression in (1). Under the null hypothesis, H_0 , the asymptotic distribution of the statistic is non standard and depends on both ρ_{xy} and c .

We do not present the limiting distribution for \mathcal{T}^* under $H_{1,W}$ because, as noted above, it depends on the distribution of s_1 . The test procedure we will subsequently develop is such that it never selects \mathcal{T}^* in large samples under Assumption W and, hence, the limit in that case is not relevant.

3.4. Asymptotic critical values for \mathcal{T}^*

Considering the case of strong dependence in Theorem 1 above, under H_0 , relevant critical values for the test based on \mathcal{T}^* will depend on the unknown nuisance parameters ρ_{xy} and c . At a practical level, and as we will show below, ρ_{xy} can be consistently estimated and so this dependence is easily dealt with (at least in large samples). The dependence on c , however, cannot be dealt with as easily because c , unlike ρ_{xy} , is not consistently estimable. We therefore adopt a scheme for simulating critical values that will, by design, yield asymptotically conservative tests. HLT propose such a method for the statistics they consider, and here we outline a similar approach for the \mathcal{T}^* statistic and its null limit distribution $\mathcal{S}^*(0, \rho_{xy}, c)$. For expository purposes we will focus attention here on upper tail critical values relevant for upper tailed tests, as this is the case of most practical relevance, but the same approach could be used in an obvious way for lower-tailed and two-tailed tests. In outlining our final preferred hybrid procedures inbreak Section 5 we will detail how to perform both upper-tailed and lower-tailed tests, and two-tailed tests.

The steps to obtaining the conservative critical values are as follows:

1. For a chosen value of ρ_{xy} , simulate the null distribution $\mathcal{S}^*(0, \rho_{xy}, c)$ for different c across an interval $c \in [0, c_{\max}]$.
2. At each value of c , compute the π -level upper-tail critical value, $cv_{\pi}^*(\rho_{xy}, c)$ say.
3. Set the π -level critical value for \mathcal{T}^* equal to $cv_{\pi}^*(\rho_{xy}) := \max_{c \in [0, c_{\max}]} cv_{\pi}^*(\rho_{xy}, c)$.

Using $cv_{\pi}^*(\rho_{xy})$ will yield a correct π -level sized test when $c = \arg \max_{c \in [0, c_{\max}]} cv_{\pi}^*(\rho_{xy}, c)$, and give a conservatively sized test for other values of c . We simulated critical values in this manner, approximating the Brownian motion processes in the limiting functional from Theorem 1 using IIDN(0, 1) random variates, and with the integrals approximated by normalized sums of 1,000 steps, with 20,000 replications (these values are used throughout our asymptotic analyses). This was carried out for the conventional significance levels $\pi \in [0.1, 0.05, 0.025, 0.01]$ for $\rho_{xy} \in [-0.95, -0.925, -0.90, \dots, 0]$, with the setting $c_{\max} = 25$ and the grid of c values being $c \in [0, 1, 2, \dots, 25]$. For these values of ρ_{xy} ,

Table 1. Response surface coefficient estimates for \mathcal{J}_{con}^* .

Regressor	$cv_{\pi}^*(\rho_{xy})$			
	$\pi = 0.1$	$\pi = 0.05$	$\pi = 0.025$	$\pi = 0.01$
1	1.275	1.673	1.957	2.332
ρ_{xy}	-1.833	0.442	-2.025	-1.980
ρ_{xy}^2	-41.401	4.969	-57.589	-52.176
ρ_{xy}^3	-466.409	3.326	-684.237	-624.069
ρ_{xy}^4	-2655.171	-17.923	-3960.352	-3657.310
ρ_{xy}^5	-8601.081	-15.799	-12872.743	-12091.819
ρ_{xy}^6	-16495.430	57.865	-24624.370	-23603.705
ρ_{xy}^7	-18541.439	53.300	-27527.328	-26985.531
ρ_{xy}^8	-11286.255	-59.652	-16640.543	-16700.284
ρ_{xy}^9	-2873.828	-62.108	-4205.605	-4321.007

$\arg \max_{c \in [0, c_{\max}]} cv_{\pi}^*(\rho_{xy}, c)$ is obtained for c much smaller than c_{\max} ; for example, with $\rho_{xy} = -0.95$, it is obtained at $c = 0$ for each value of π .

To automate selection of an appropriate critical value for a given value of ρ_{xy} , we calculated a response surface by regressing $cv_{\pi}^*(z)$ on $F(z) := [1, z, z^2, \dots, z^9]$ with $z = \rho_{xy}$ for the 38 data points corresponding to the grid of values for ρ_{xy} .² The response surface critical value is the fitted value from this regression, and the coefficient estimates are given in Table 1. In practice, the response surface critical values can be calculated by substituting the unknown correlation parameter ρ_{xy} with a consistent estimate. To that end, as in HLT, we suggest using the estimator

$$\hat{\rho}_{xy} := \frac{\sum_{t=2}^T \hat{\epsilon}_{xt} \hat{\epsilon}_{yt}}{\sqrt{\sum_{t=2}^T \hat{\epsilon}_{xt}^2 \sum_{t=2}^T \hat{\epsilon}_{yt}^2}} \quad (8)$$

where $\hat{\epsilon}_{yt}$ are the OLS residuals from regressing y_t on a constant and x_{t-1} , and where it is recalled from Section 3.2 that $\hat{\epsilon}_{xt}$ denote the OLS residuals from regressing Δx_t on a constant and x_{t-1} . It is straightforward to show that $\hat{\rho}_{xy}$ is a consistent estimator of ρ_{xy} under either Assumption S or Assumption W. In what follows, we denote tests based on comparison of \mathcal{J}^* with an asymptotically conservative critical value by \mathcal{J}_{con}^* .

3.5. Alternative feasible tests

One alternative feasible test is the standard t -statistic \mathcal{J} . Under Assumption W, it is straightforward to show that \mathcal{J} has a standard normal limiting null distribution for any value of ρ_{xy} , and thus has the potential for nuisance parameter free inference in this world. With respect to the DGP in (4), \mathcal{J} is based on a correctly specified regression when $\rho_{xy} = 0$, but when $\rho_{xy} \neq 0$, the regression omits a relevant regressor; while this does not affect the limiting null distribution, \mathcal{J} will be inefficient relative to the infeasible test if $\rho_{xy} \neq 0$. However, among feasible tests, \mathcal{J} is asymptotically optimal (under Gaussianity) for all ρ_{xy} (see Jansson and Moreira, 2006, p.704), hence we would wish to apply this test under Assumption W, as is done in HLT's hybrid procedure. Theorem 2 of HLT shows that under Assumption W, as $T \rightarrow \infty$, $\mathcal{J} \Rightarrow N(g, 1)$ under $H_{1,W}$. Under Assumption S, \mathcal{J} has a standard normal limit null distribution provided $\rho_{xy} = 0$, in which case it is also efficient; whenever $\rho_{xy} \neq 0$, however, its limit null distribution depends on ρ_{xy} and c .

²The response surface functional form $F(z)$ was adopted after considerable experimentation and is similar to that employed in HLT, but with the addition of a ρ_{xy}^9 term and a finer grid of ρ_{xy} values. The response surface was found to work very well, with the regression R^2 being in excess of 0.999.

A second feasible statistic proposed by HLT is a variant of the standard t -statistic, appropriate in the case of strongly persistent x_t , taking the form of the t -statistic associated with the OLS estimate of β in the regression

$$(y_t - \hat{\alpha}_y) = \beta(x_{t-1} - \tilde{\alpha}_x) + v_t \quad (9)$$

where $\hat{\alpha}_y := (T-1)^{-1} \sum_{t=2}^T y_t$. We denote this statistic as \mathcal{T}' . Under [Assumption W](#), the limiting null distribution of \mathcal{T}' will depend on the (unknown) distribution of s_1 (as with \mathcal{T}^*), hence the statistic is again only designed for use in the strongly persistent world (in contrast to \mathcal{T}). Under [Assumption S](#), it follows from [Theorem 1](#) of HLT that

$$\begin{aligned} \mathcal{T} &\Rightarrow \frac{g\sigma_x}{\sigma_y} \sqrt{\int_0^1 \bar{W}_{1c}(r)^2 dr} + \frac{\int_0^1 \bar{W}_{1c}(r) d\left\{\rho_{xy} W_1(r) + \sqrt{1 - \rho_{xy}^2} W_2(r)\right\}}{\sqrt{\int_0^1 \bar{W}_{1c}(r)^2 dr}} =: \mathcal{S}(g\sigma_x/\sigma_y, \rho_{xy}, c) \\ \mathcal{T}' &\Rightarrow \frac{g\sigma_x}{\sigma_y} \frac{\int_0^1 \bar{W}_{1c}(r)^2 dr}{\sqrt{\int_0^1 W_{1c}(r)^2 dr}} + \frac{\int_0^1 \bar{W}_{1c}(r) d\left\{\rho_{xy} W_1(r) + \sqrt{1 - \rho_{xy}^2} W_2(r)\right\}}{\sqrt{\int_0^1 W_{1c}(r)^2 dr}} =: \mathcal{S}'(g\sigma_x/\sigma_y, \rho_{xy}, c). \end{aligned}$$

In what follows, we denote tests based on comparison of \mathcal{T} and \mathcal{T}' with asymptotically conservative critical values by \mathcal{T}_{con} and \mathcal{T}'_{con} , respectively (response surfaces for the conservative critical values are provided in HLT).

3.6. Asymptotic local power comparisons under strong persistence

Under [Assumption S](#), we can use the limiting representations given in [Theorem 1](#) to compare the asymptotic local powers of tests based on the \mathcal{T} , \mathcal{T}' and \mathcal{T}^* statistics for a range of values of the relevant nuisance parameters on which these depend, ρ_{xy} and c .³ We simulate $\mathcal{S}(g, \rho_{xy}, c)$, $\mathcal{S}'(g\sigma_x/\sigma_y, \rho_{xy}, c)$ and $\mathcal{S}^*(g\sigma_x/\sigma_y, \rho_{xy}, c)$ and compare these to the relevant conservative critical values. In what follows, we set $\pi = 0.05$ and conduct upper tail tests. For a given value of ρ_{xy} and c , we compute asymptotic powers across $g \geq 0$ ($g = 0$ representing asymptotic size). We consider $\rho_{xy} \in [-0.95, -0.7, -0.5, -0.1]$ and $c \in [0, 1.25, 2.5, 5, 10, 25, 50, 100]$. For positive values of ρ_{xy} , we find a result similar to HLT in that \mathcal{T}_{con} becomes the best performing test as ρ_{xy} increases; in the hybrid procedure that we later propose, we follow HLT and make use of \mathcal{T}_{con} for $\hat{\rho}_{xy} > -0.1$, hence here our focus is on negative values of ρ_{xy} . Note that this is also the reason why our response surfaces for \mathcal{T}^*_{con} outlined above were based only on non positive values of ρ_{xy} .

The results for $\rho_{xy} = -0.95$ are given in [Fig. 1](#). For $c = 0$ we see that \mathcal{T}^*_{con} is more powerful than \mathcal{T}_{con} and \mathcal{T}'_{con} , substantially so with respect to \mathcal{T}'_{con} . This remains the case for $c = 1.25$ and $c = 2.5$; we observe the power advantage over \mathcal{T}_{con} increasing for these values of c , although the power advantage of \mathcal{T}^*_{con} relative to \mathcal{T}'_{con} is diminishing as c increases. Once $c = 5$ or greater, \mathcal{T}^*_{con} is only marginally more powerful than \mathcal{T}'_{con} , but both are considerably more powerful than \mathcal{T}_{con} . In [Figs. 2–4](#), the analysis is repeated with $\rho_{xy} = -0.7$, $\rho_{xy} = -0.5$, and $\rho_{xy} = -0.1$. Again \mathcal{T}^*_{con} is more powerful than \mathcal{T}'_{con} for the lower values of c and they appear to be very similar for the higher values of c . Comparing \mathcal{T}^*_{con} with \mathcal{T}_{con} , we observe similar patterns of relative power behavior in [Figs. 2](#) and [3](#) as were seen in [Fig. 1](#), with \mathcal{T}^*_{con} outperforming \mathcal{T}_{con} , increasingly so as c increases. However, as ρ_{xy} becomes less negative, the differences in power between \mathcal{T}^*_{con} and \mathcal{T}_{con} become less marked. Indeed, in [Fig. 4](#) where $\rho_{xy} = -0.1$, the powers of \mathcal{T}^*_{con} and \mathcal{T}_{con} become almost indistinguishable across almost all c .

³Recall from [Remark 1](#) that these limiting distributions do not depend on any nuisance parameters arising from conditional heteroskedasticity in the innovations of the form given in [Assumption A.1](#) of [Campbell and Yogo \(2006\)](#), and so the analysis which follows remains valid in the presence of conditional heteroskedasticity.

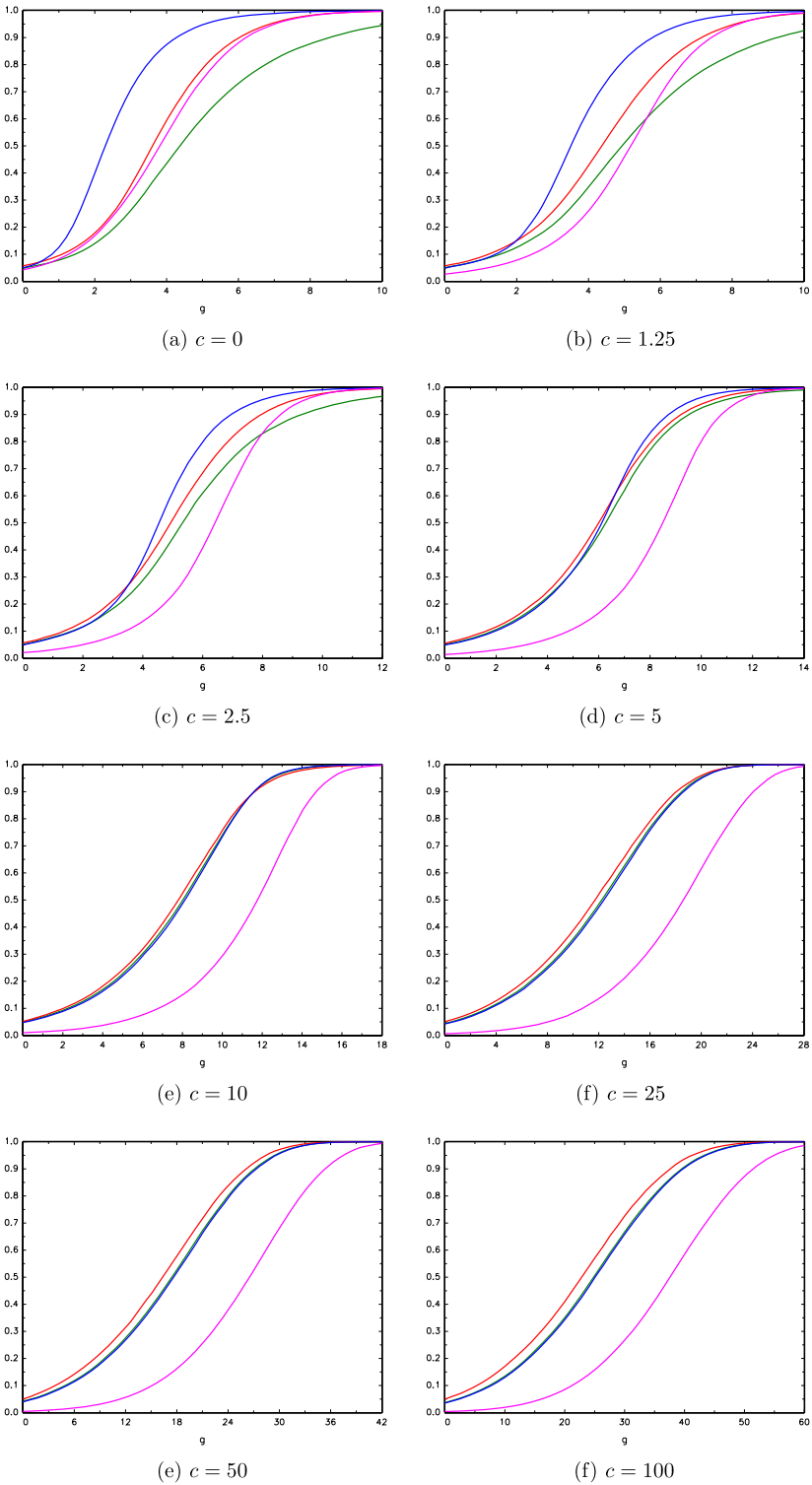


Figure 1. Asymptotic local power of nominal 0.05-level tests, $\rho_{xy} = -0.95$; \mathcal{J}_{con}^* : —, \mathcal{J}'_{con} : —, \mathcal{J}_{con} : —, \mathcal{J}_{con}^W : —.

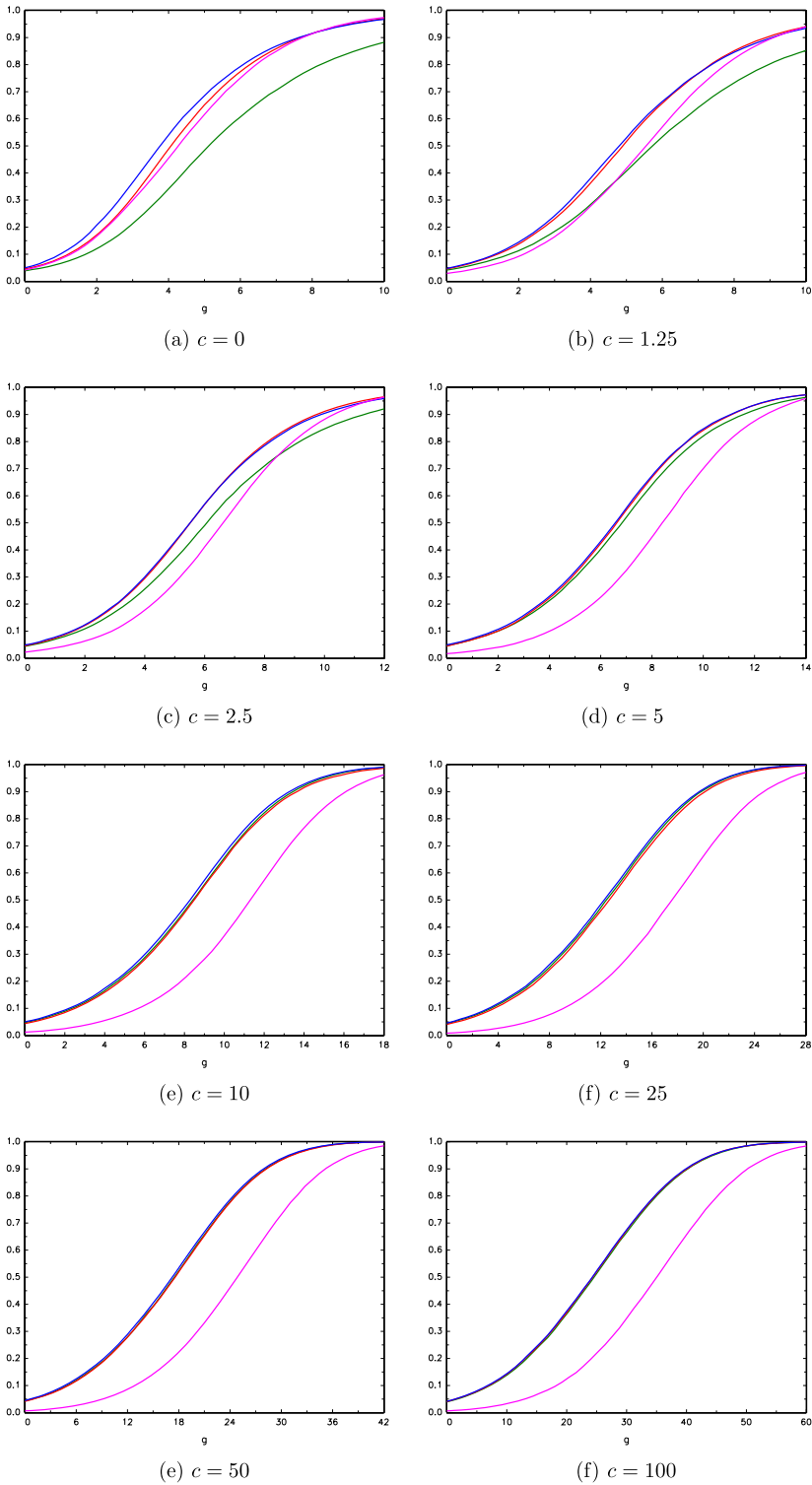


Figure 2. Asymptotic local power of nominal 0.05-level tests, $\rho_{xy} = -0.7$; \mathcal{J}'_{con} : —, \mathcal{J}'_{con} : —, \mathcal{J}_{con} : —, \mathcal{J}^W_{con} : —.

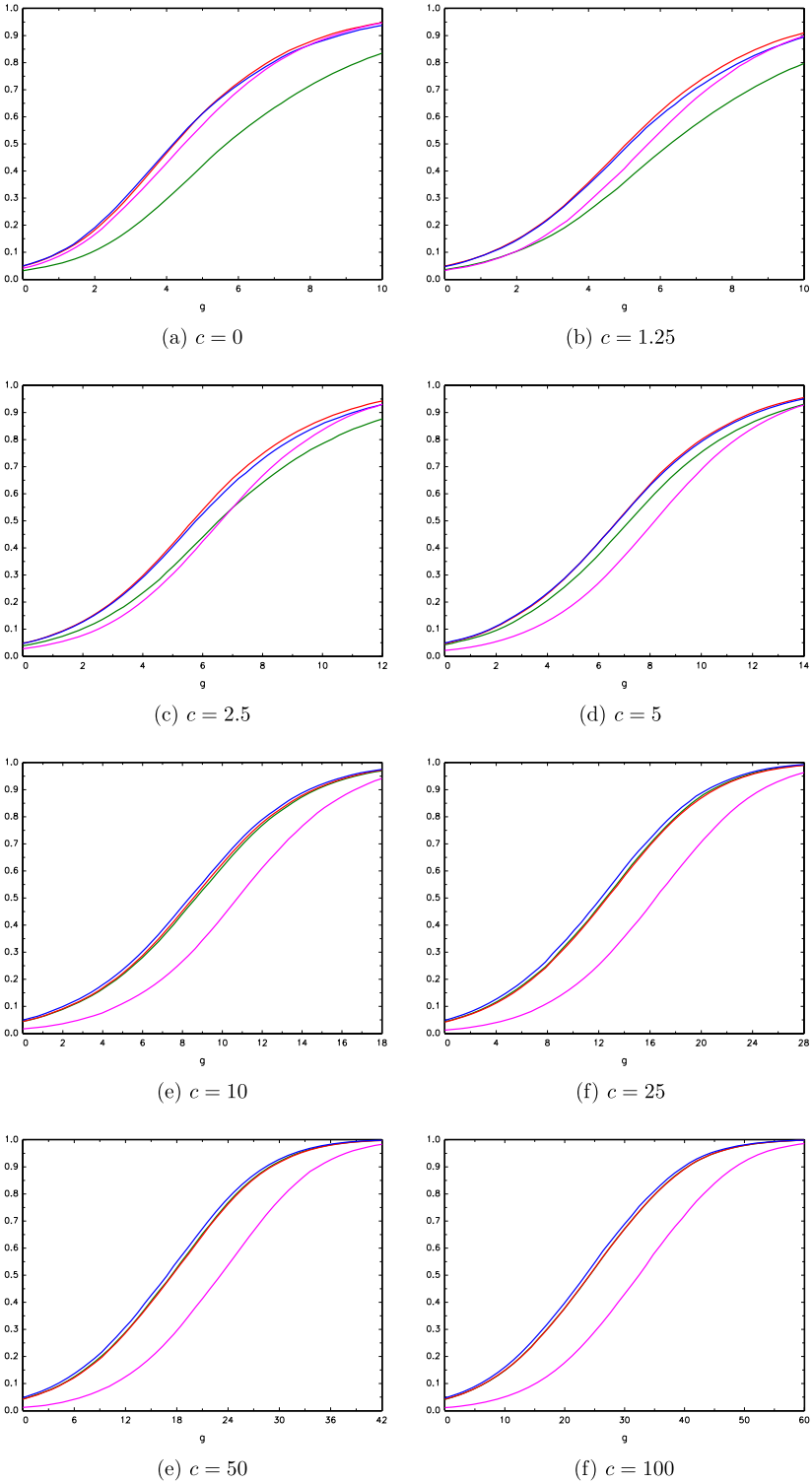


Figure 3. Asymptotic local power of nominal 0.05-level tests, $\rho_{xy} = -0.5$; $\mathcal{J}'_{con}*$: —, \mathcal{J}'_{con} : —, \mathcal{J}_{con} : —, \mathcal{J}^W_{con} : —

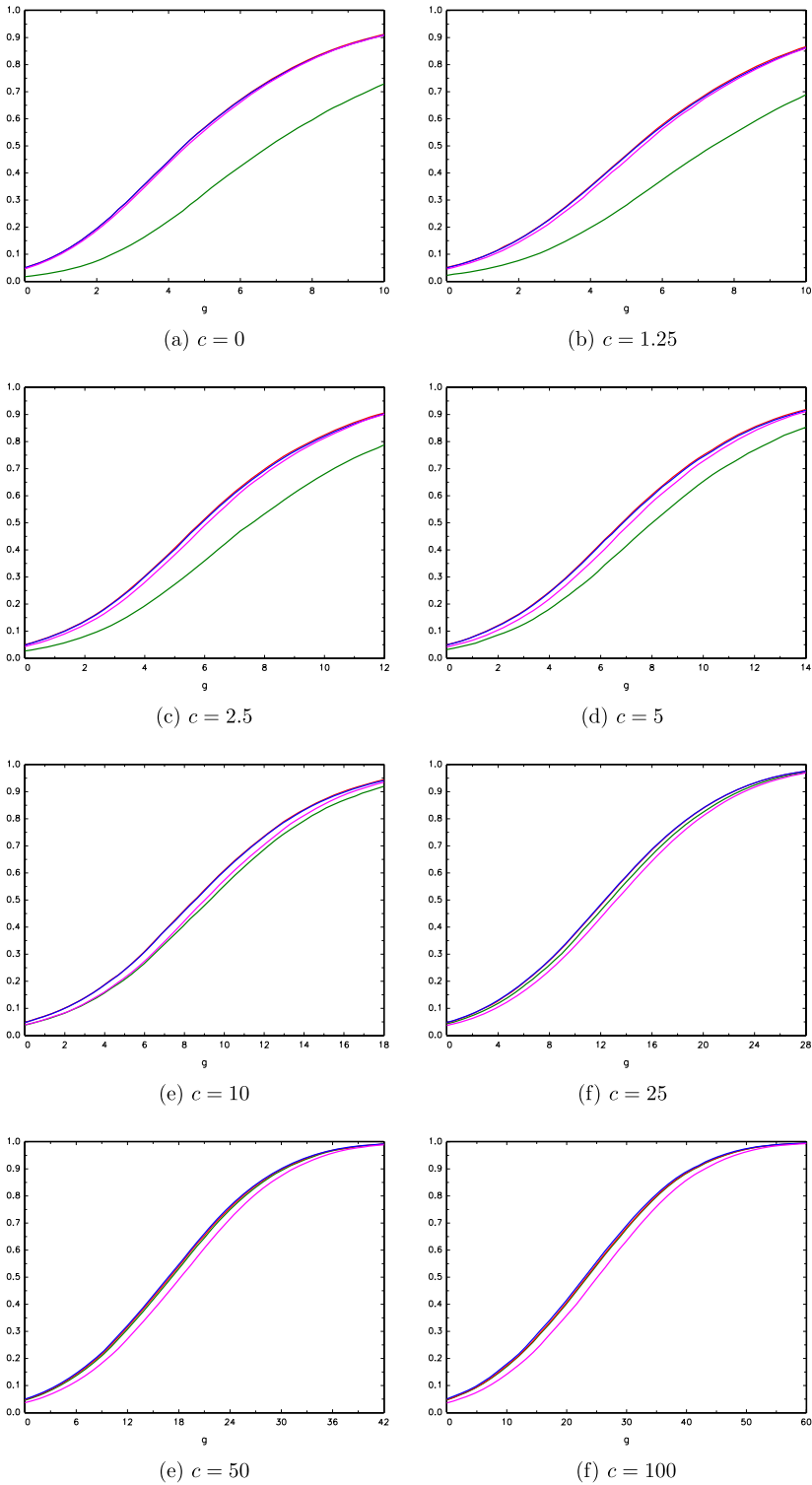


Figure 4. Asymptotic local power of nominal 0.05-level tests, $\rho_{xy} = -0.1$; \mathcal{J}'_{con^*} : —, \mathcal{J}'_{con} : —, \mathcal{J}^W_{con} : —, \mathcal{J}^W_{con} : —

4. A weighted test under strong persistence

Given the asymptotic power simulation results reported above, it is interesting to consider whether we might be able to combine \mathcal{T}^* and \mathcal{T}' (the two best performing tests) in a way to possibly improve power over and above that displayed by \mathcal{T}^* . For the purposes of illustration, our arguments will concentrate on the environment of Fig. 1(a), where $\rho_{xy} = -0.95$ and $c = 0$ where, as noted above, \mathcal{T}^* is clearly the more powerful test. Now, under H_0 the correlation between \mathcal{T}^* and \mathcal{T}' is 0.90. A consequence of this high level of correlation is that the rejections obtained from \mathcal{T}' under $H_{1,S}$ are close to being a subset of those obtained from \mathcal{T}^* . This implies that any (linear) combination of \mathcal{T}^* and \mathcal{T}' of the form $w_f^* \mathcal{T}^* + w_f' \mathcal{T}'$, where w_f^* and w_f' are some fixed positive weights standardized such that $w_f^* + w_f' = 1$, cannot lead to improved levels of power above that of \mathcal{T}^* because the correlation between the components $w_f^* \mathcal{T}^*$ and $w_f' \mathcal{T}'$ is identical to that between \mathcal{T}^* and \mathcal{T}' . We can, however, consider a *randomized* weighting scheme, with the random weights w_r^* and w_r' , say, determined from the available data, and with w_r^* and w_r' having support on $[0, 1]$ and $w_r^* + w_r' = 1$. The aim is for $w_r^* \mathcal{T}^*$ and $w_r' \mathcal{T}'$ to have lower correlation than holds between \mathcal{T}^* and \mathcal{T}' , although it is crucial that the two components remain positively correlated. In this way, each component can potentially make a greater individual contribution to overall power. In what follows we will use x_t to construct the weights because, unlike y_t , its behavior does not depend on whether H_0 or $H_{1,S}$ is true.

The distribution of any data dependent weights based on x_t will, of course, depend on the value of c . The weight scheme we consider here is based on the p -value, denoted p^{NB} , associated with the familiar local-GLS demeaned normalized bias unit root test statistic of Elliott et al. (1996), i.e., $NB := T\tilde{\phi}$ in the context of (6). Well known results show that, under Assumption S, $NB \Rightarrow (\int_0^1 W_{1c}(r) dW_{1c}(r)) / (\int_0^1 W_{1c}(r)^2 dr) =: \mathcal{S}^{NB}(c)$, and denoting the density function of $\mathcal{S}^{NB}(0)$ by $f(x)$, we can write $p^{NB} \Rightarrow \int_{-\infty}^{\mathcal{S}^{NB}(c)} f(x) dx =: p^{NB}(\mathcal{S}^{NB}(c))$. The attractive feature of using p^{NB} in the weight scheme is that when $c = 0$, $p^{NB} \Rightarrow p^{NB}(\mathcal{S}^{NB}(0)) = U(0, 1)$. Hence in the $c = 0$ case where most difference is observed between the power profiles of \mathcal{T}^* and \mathcal{T}' , the weights can be based on a uniformly distributed variate on $(0, 1)$. As c becomes large, the power gains of \mathcal{T}^* over \mathcal{T}' diminish and consequently the role of the weight function becomes less critical; use of a weight scheme based on p^{NB} is again appealing here, because as $c \rightarrow \infty$, $p^{NB} \Rightarrow 0$.

Following such considerations, the weights we consider are defined as $w_r^* := (p^{NB})^\lambda$ and $w_r' := 1 - (p^{NB})^\lambda$, where the positive constant λ is introduced to permit an additional degree of calibration in the weight specification. We therefore consider the weighted statistic

$$\mathcal{T}^w := w_r^* \mathcal{T}^* + w_r' \mathcal{T}'.$$

With this weighting scheme, the asymptotic correlation between $w_r^* \mathcal{T}^*$ and $w_r' \mathcal{T}'$ remains positive across all values of ρ_{xy} and c we consider (and for all λ). The weighted statistic \mathcal{T}^w is thus comprised of a weighted average of \mathcal{T}^* and \mathcal{T}' , with the weighted average of the tests having most effect when $c = 0$ and reducing towards simply \mathcal{T}' as c becomes large.

In Theorem 2 we next state the limiting distribution of \mathcal{T}^w under Assumption S. The stated result follows straightforwardly from the result given in Theorem 1 and the limiting distribution for \mathcal{T}' given in Section 3.5.

Theorem 2. *Under the conditions of Theorem 1,*

$$\begin{aligned} \mathcal{T}^w &\Rightarrow \{p^{NB}(\mathcal{S}^{NB}(c))\}^\lambda \mathcal{S}^*(g\sigma_x/\sigma_y, \rho_{xy}, c) + [1 - \{p^{NB}(\mathcal{S}^{NB}(c))\}^\lambda] \mathcal{S}'(g\sigma_x/\sigma_y, \rho_{xy}, c) \\ &=: \mathcal{S}_{NB}^w(g\sigma_x/\sigma_y, \rho_{xy}, c, \lambda). \end{aligned}$$

In order to implement \mathcal{T}^w , we obtain a response surface for p^{NB} based on simulated limit distributions. We simulated the limit of NB under $c = 0$, i.e., $\mathcal{S}^{NB}(0)$, and then calculated the numerical approximation to $p^{NB}(x)$ for $x \in [-20, -19.95, -19.9, \dots, 4]$. To automate selection of an appropriate asymptotic p -

Table 2. Response surface coefficient estimates for p^{NB} .

Regressor	$p^{NB}(x)$
1	0.0004
$z^{0.25}$	0.2706
$z^{0.5}$	0.6951
z	-0.8366
z^2	2.2851
z^3	-1.4128

Note: $z = 1/(1 + e^{-x})$

Table 3. Response surface coefficient estimates for \mathcal{T}_{con}^w .

Regressor	$\lambda_{max}(\rho_{xy})$				$cv_{\pi}^*(\rho_{xy})$			
	$\pi = 0.1$	$\pi = 0.05$	$\pi = 0.025$	$\pi = 0.01$	$\pi = 0.1$	$\pi = 0.05$	$\pi = 0.025$	$\pi = 0.01$
1	0.061	0.050	0.098	0.043	1.264	1.625	1.918	2.308
ρ_{xy}	-3.724	0.088	0.601	-0.315	1.517	-0.955	-0.314	-0.506
ρ_{xy}^2	16.242	-23.567	-9.284	8.438	28.840	-14.187	10.874	-9.799
ρ_{xy}^3	479.214	-627.685	-161.781	187.413	223.885	-87.859	147.135	-97.115
ρ_{xy}^4	3418.236	-4773.695	-890.161	1811.938	985.886	-234.740	878.957	-501.689
ρ_{xy}^5	12052.607	-17736.573	-2701.112	7722.234	2578.918	-221.424	2875.649	-1483.912
ρ_{xy}^6	23830.004	-36158.566	-4747.985	16982.123	4081.238	224.125	5448.789	-2547.421
ρ_{xy}^7	26653.271	-41265.834	-4830.732	20165.939	3832.415	705.946	5947.498	-2476.120
ρ_{xy}^8	15696.413	-24739.360	-2667.365	12281.453	1964.964	596.176	3468.018	-1245.259
ρ_{xy}^9	3777.182	-6067.789	-623.772	3006.942	423.771	175.342	836.452	-246.514

value for a given value of x , we once again calculated a response surface by regressing $p^{NB}(x)$ on $G(z) := [1, z^{0.25}, z^{0.5}, z, z^2, z^3]$ with $z = 1/(1 + e^{-x})$ (481 data points), the logistic function z being a natural choice given that we are approximating a cumulative density function.⁴ The response surface p -value is the fitted value from this regression, and the response surface coefficient estimates are provided in Table 2, denoted $p^{NB}(x)$. In practice, the response surface p -value can be calculated using $x = NB$.

4.1. Asymptotic critical values and selection of λ

Calculation of conservative asymptotic critical values for the \mathcal{T}^w statistic is carried out in exactly the same manner as for the \mathcal{T}^* statistic in Section 3.4, but based on $S_{NB}^w(0, \rho_{xy}, c, \lambda)$. For a given value of λ , we obtain the conservative critical value $cv_{\pi}^w(\rho_{xy}, \lambda)$. At the same time, we evaluate the local alternative distribution $S_{NB}^w(g\sigma_x/\sigma_y, \rho_{xy}, c, \lambda)$ with $\sigma_x = \sigma_y = 1$ for $g = 7.5$ over $c \in [0, 1, 2, \dots, 25]$ and compare this with $cv_{\pi}^w(\rho_{xy}, \lambda)$. We choose λ to maximize the average power across c , where the candidate values of λ we consider are $\lambda \in [0.05, 0.1, 0.15, \dots, 2.5]$. We denote the power-maximizing value of λ as $\lambda_{max}(\rho_{xy})$ and the corresponding conservative critical value as $cv_{\pi}^w(\rho_{xy})$. As with $cv_{\pi}^*(\rho_{xy})$ in the context of \mathcal{T}_{con}^* , we found $cv_{\pi}^w(\rho_{xy})$ is obtained with $c = 0$ when $\rho_{xy} = -0.95$.

To select the appropriate value of λ and conservative critical value for a given value of ρ_{xy} , we calculated response surfaces by regressing $\lambda_{max}(\rho_{xy})$ and $cv_{\pi}^w(\rho_{xy})$, respectively, once again on $F(z) = [1, z, z^2, \dots, z^9]$ with $z = \rho_{xy}$ for the 38 data points in our grid of values for ρ_{xy} . The response surface coefficient estimates for the ρ_{xy} -dependent values of λ and associated critical values for $\pi \in [0.10, 0.05, 0.025, 0.01]$ can be found in Table 3 (the remarks made in footnote 2 apply here also). We will refer to this testing procedure in what follows as \mathcal{T}_{con}^w .

⁴The response surface functional form $G(z)$ was adopted after considerable experimentation and differs from that used for $F(z)$ in Section 3.4 due to the use of the logistic function and p -values, with relatively few terms required for a good fit. The response surface was found to work very well, with the regression R^2 again exceeding 0.999.

4.2. Asymptotic local power comparisons of \mathcal{T}^* , \mathcal{T}' and \mathcal{T}^w under strong persistence

Figure 1(a) also presents simulations of the asymptotic powers of the \mathcal{T}_{con}^w procedure for $c = 0$, again using upper tail tests for $\pi = 0.05$, eliciting a direct comparison with \mathcal{T}_{con}^* and \mathcal{T}'_{con} . We see that \mathcal{T}_{con}^w (calculated using $\lambda_{\max}(\rho_{xy})$ and compared to its conservative critical value $cn_{\pi}^w(\rho_{xy})$) is substantially more powerful than \mathcal{T}_{con}^* . So we obtain a situation where combining the tests produces useful gains in the sense that the combined procedure has higher power than both of the individual constituent tests. This is made possible because the components of \mathcal{T}^w , i.e., $(p^{NB})^\lambda \mathcal{T}^*$ and $\{1 - (p^{NB})^\lambda\} \mathcal{T}'$, have asymptotic correlation 0.43 under H_0 , which is positive but much lower than that of \mathcal{T}^* and \mathcal{T}' (0.90). It is also interesting to note that the critical value of \mathcal{T}_{con}^w here is 1.96, which is close to the critical value of \mathcal{T}'_{con} (1.94) and substantially smaller than that of \mathcal{T}_{con}^* (5.40).

For the other values of c in Fig. 1, we see \mathcal{T}_{con}^w still dominating \mathcal{T}_{con}^* (and hence \mathcal{T}'_{con}) until $c = 10$. At this point its power essentially coincides with that of \mathcal{T}'_{con} since p^{NB} is now generally close to zero. For $\rho_{xy} = -0.7$ and $\rho_{xy} = -0.5$ in Figs. 2 and 3, respectively, we see that the power levels of \mathcal{T}_{con}^w are near to those of \mathcal{T}_{con}^* , even for small values of c where \mathcal{T}_{con}^* is more powerful than \mathcal{T}'_{con} . Hence it is not always the case that the weighted combination improves upon the better of \mathcal{T}_{con}^* and \mathcal{T}'_{con} , but we do find that \mathcal{T}_{con}^w is never meaningfully outperformed by the better of the two individual tests. When $\rho_{xy} = -0.1$ in Fig. 4, \mathcal{T}_{con}^w has a power profile that essentially coincides with \mathcal{T}_{con}^* .

We investigated the effects of switching the weights in \mathcal{T}^w such that $w_r^* = 1 - (p^{NB})^\lambda$ and $w_r' = (p^{NB})^\lambda$ in the case of $\rho_{xy} = -0.95$ and $c = 0$ (i.e., the settings of Fig. 1(a)). The components of this variant of \mathcal{T}^w , $\{1 - (p^{NB})^\lambda\} \mathcal{T}^*$ and $(p^{NB})^\lambda \mathcal{T}'$, now have asymptotic correlation -0.38 under H_0 and the critical value of \mathcal{T}_{con}^w is 5.66, somewhat larger than that of \mathcal{T}_{con}^* (5.40). The powers of \mathcal{T}_{con}^w were found to be uniformly below those of \mathcal{T}'_{con} , let alone \mathcal{T}_{con}^* , which serves to illustrate the importance of the components of \mathcal{T}^w being positively correlated.

5. A hybrid procedure allowing for strong or weak persistence

Although the main focus of our analysis thus far has been on the case of strong persistence, we now outline our proposed hybrid testing procedure which closely mirrors the hybrid testing procedure, denoted \mathcal{T}_{hyb} in what follows, outlined in Section 3.3 of HLT. This procedure is designed to capitalize on the optimality property of the conventional t -test (where \mathcal{T} is compared to a standard normal critical value) under weak persistence (Assumption W), and exploit the relative local power advantages of \mathcal{T}_{con}^w and \mathcal{T}_{con} observed from the analysis in Section 4.2 for different values of ρ_{xy} under strong persistence (Assumption S). This will entail the use of two switching mechanisms. The first involves a switching approach similar to that used in EMW, whereby the standard test is selected when evidence of a weakly persistent predictor is present. In the absence of such evidence, a secondary switching mechanism is needed to determine whether \mathcal{T}_{con}^w or \mathcal{T}_{con} should be applied, this time on the basis of a consistent estimate of ρ_{xy} ; in particular, for a strongly persistent predictor we would want to make use of \mathcal{T}_{con}^w for more negative values of ρ_{xy} , and \mathcal{T}_{con} for small negative and positive ρ_{xy} .

The hybrid testing procedure we outline below can therefore be seen to parallel the structure of the \mathcal{T}_{hyb} procedure of HLT, with the statistic \mathcal{T}' in HLT's procedure replaced by the weighted statistic \mathcal{T}^w , developed in Section 4 above, in the light of its superior power performance documented in Section 4.2. Denoting such a procedure by \mathcal{T}_{hyb}^w , our proposed hybrid testing approach proceeds as follows:

1. If $NB_{OLS} < -4T^{1/2}$ perform \mathcal{T}_N , where \mathcal{T}_N denotes the test which compares \mathcal{T} with a standard normal critical value, and where $NB_{OLS} := T\hat{\phi}$ is the standard OLS demeaned Dickey-Fuller normalized bias unit root statistic based on $\hat{\phi}$, the OLS slope estimate obtained from regressing Δx_t on a constant and x_{t-1} .

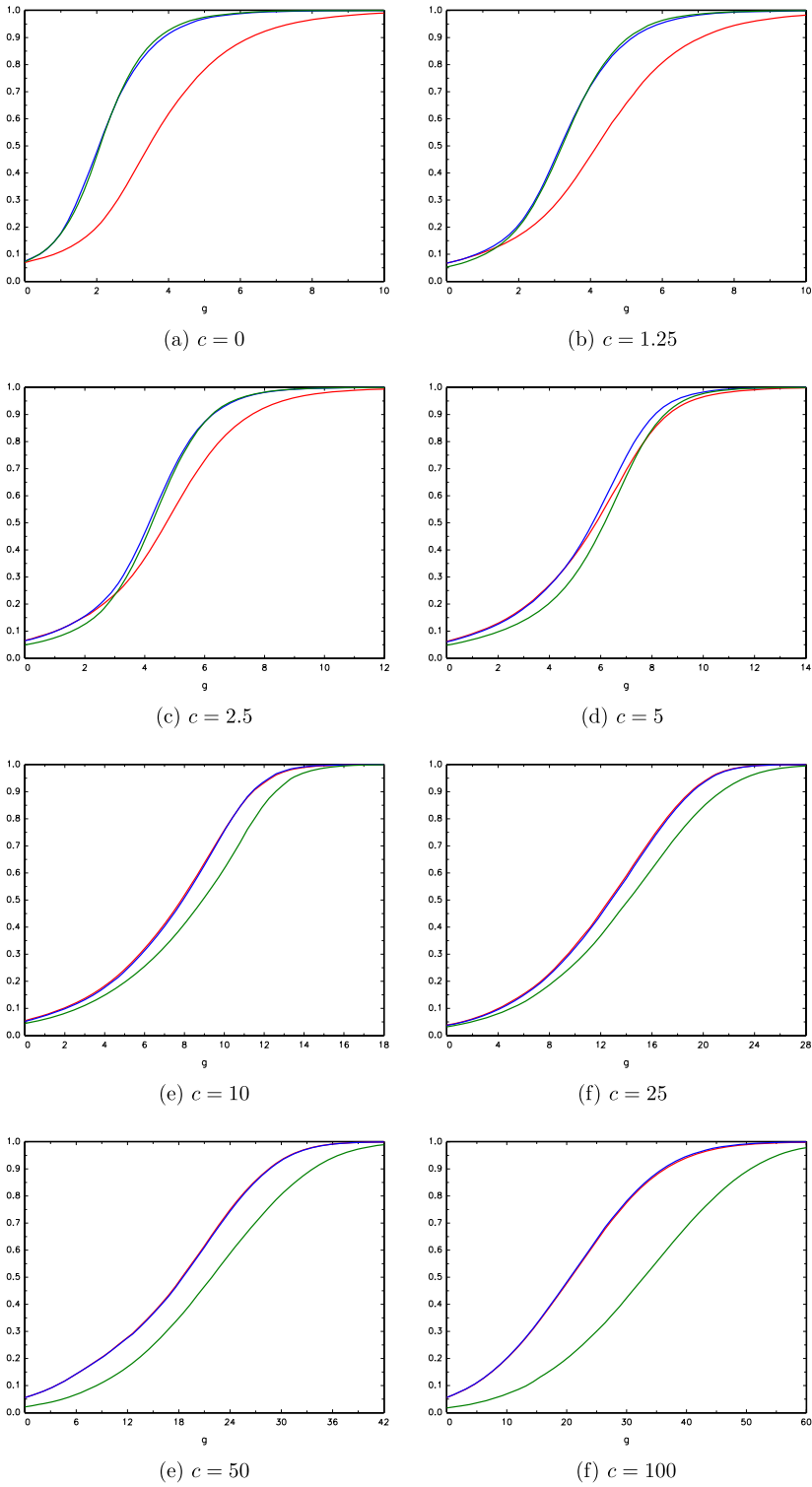


Figure 5. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.95$; J_{hyb}^W : —, EMW : —, J_{hyb} : —.

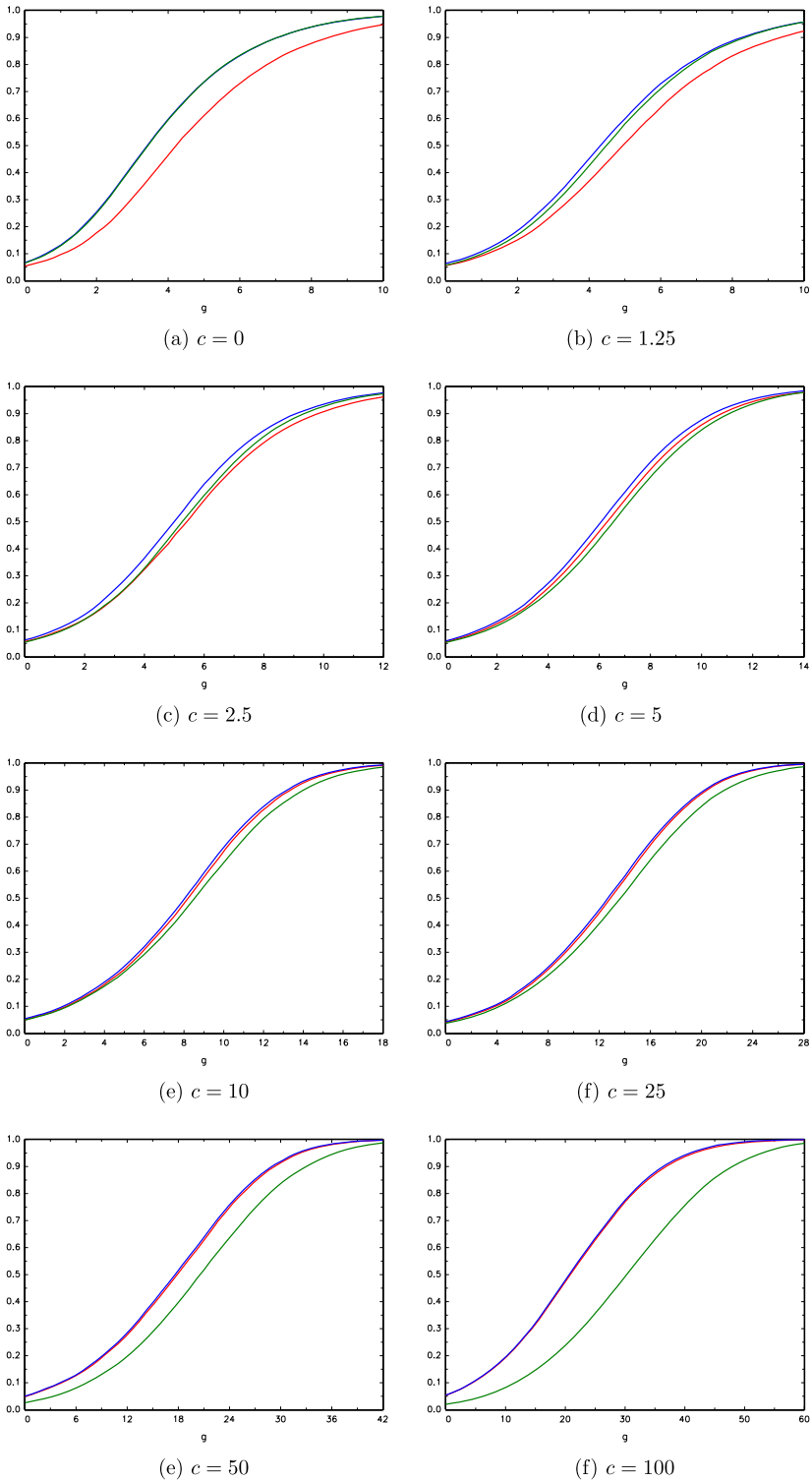


Figure 6. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.7$; \mathcal{J}_{hyb} : —, \mathcal{J}_{hyb}^W : —, EMW : —.

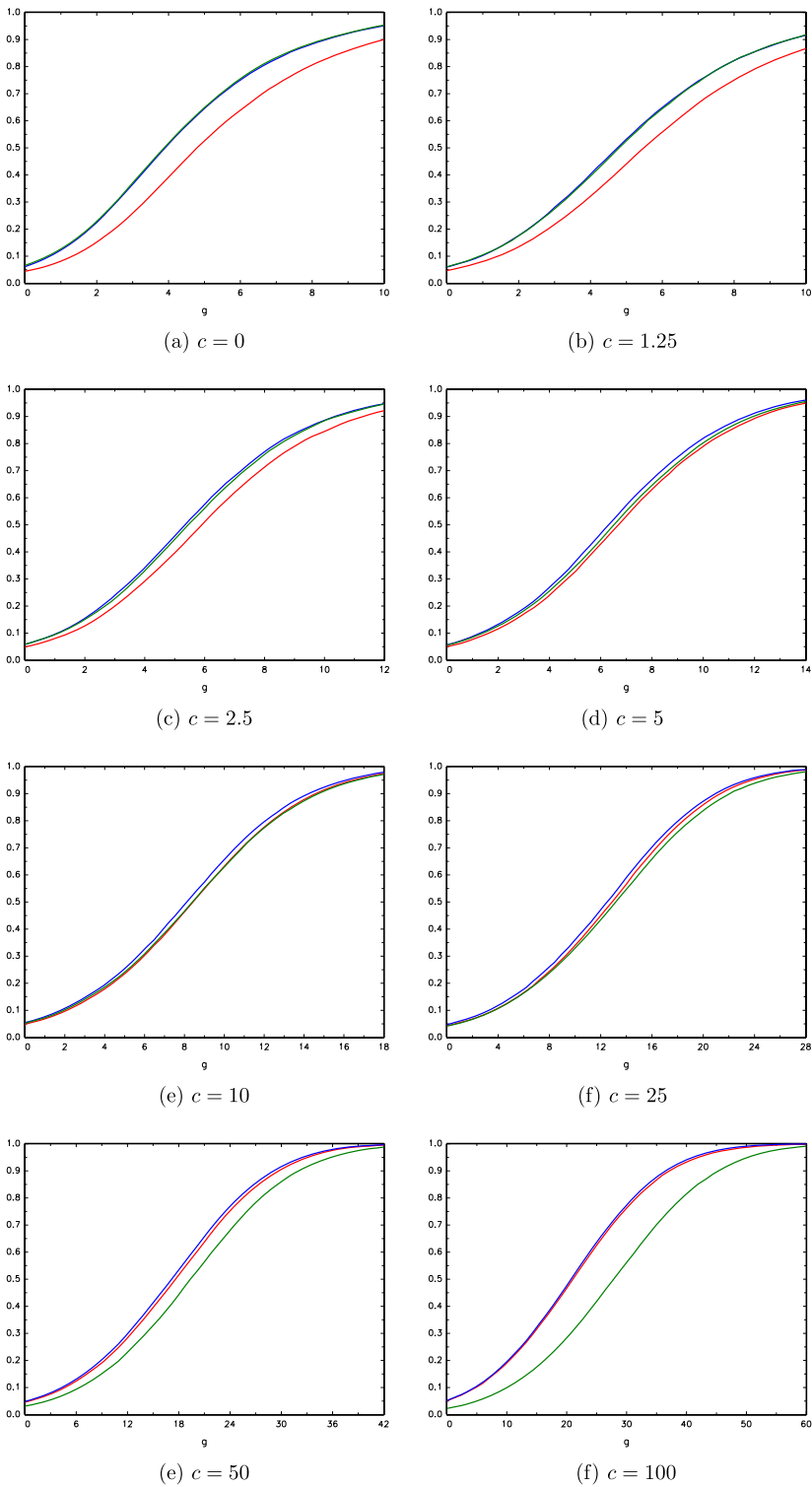


Figure 7. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.5$; \mathcal{J}_{hyb} : —, \mathcal{J}_{hyb}^W : —, EMW : —.

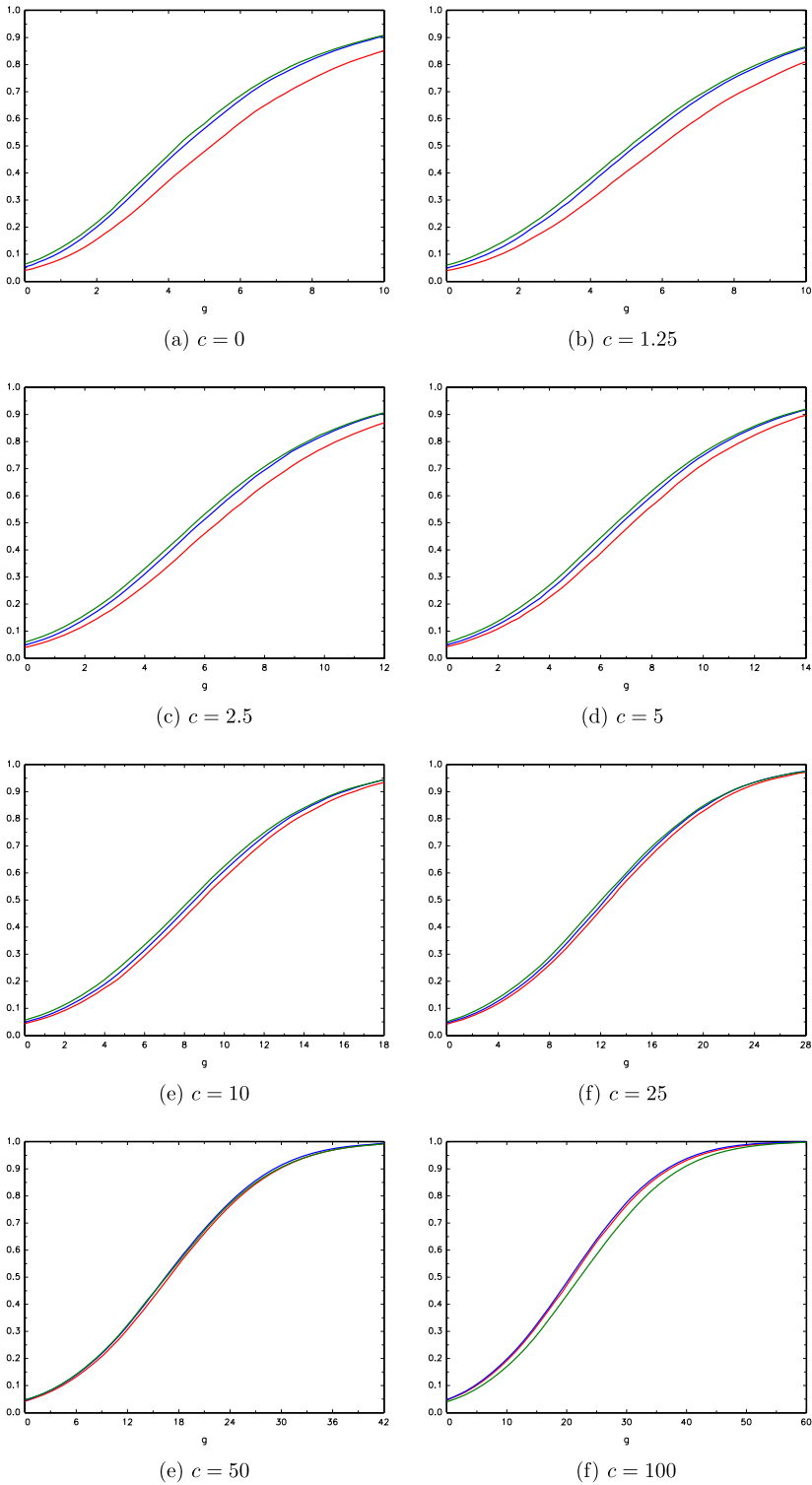


Figure 8. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.1$; \mathcal{J}_{hyb} : —, \mathcal{J}_{hyb}^W : —, EMW : —.

2. Otherwise:

(a) For upper-tail tests against the alternative $\beta > 0$,

if $\hat{\rho}_{xy} > -0.1$ perform \mathcal{J}_{con} (\mathcal{J} with a conservative critical value)
 if $\hat{\rho}_{xy} < -0.1$ perform \mathcal{J}_{con}^w (\mathcal{J}^w with $\lambda_{\max}(\hat{\rho}_{xy})$ and conservative critical value $cv_{\pi}^w(\hat{\rho}_{xy})$)

(b) For lower-tail tests against the alternative $\beta < 0$,

if $\hat{\rho}_{xy} < 0.1$ perform \mathcal{J}_{con} (\mathcal{J} with a conservative critical value)
 if $\hat{\rho}_{xy} > 0.1$ perform \mathcal{J}_{con}^w (\mathcal{J}^w with $\lambda_{\max}(-\hat{\rho}_{xy})$ and conservative critical value $-cv_{\pi}^w(-\hat{\rho}_{xy})$)

Remark 4. Step 1 coincides with Step 1 of the corresponding hybrid testing procedure, \mathcal{J}_{hyb} , from HLT. As in HLT, the normalized bias statistic is used to distinguish between the strongly and weakly persistent cases. Under [Assumption S](#), $NB_{OLS} = O_p(1)$, while under [Assumption W](#), NB_{OLS} diverges to minus infinity. For the reasons outlined on p.205 of HLT, we implement NB_{OLS} with a sample size dependent critical value of $-4T^{1/2}$. Under [Assumption W](#), NB_{OLS} diverges to infinity at a rate faster than $T^{1/2}$, hence \mathcal{J}_N is always selected asymptotically under weak persistence because $Pr(NB_{OLS} < -4T^{1/2}) \rightarrow 1$ as the sample size diverges.

Remark 5. Under strong persistence and the values of ρ_{xy} we have considered in our asymptotic power analysis in [Section 4.2](#), the asymptotic behavior of the hybrid test procedure \mathcal{J}_{hyb} of HLT coincides with that of \mathcal{J}'_{con} , while the asymptotic behavior of the new procedure \mathcal{J}_{hyb}^w coincides with that of \mathcal{J}_{con}^w . As such, \mathcal{J}_{hyb}^w will have considerably higher asymptotic power than \mathcal{J}_{hyb} in this environment when c is small.

Remark 6. Although we have outlined our hybrid testing procedure in terms of upper- and lower-tail one-sided tests for predictability, in principle these could also be used to perform two sided tests for predictability. In particular, supposing the upper- and lower-tail versions of the test were both run at the (asymptotic) $\pi/2\%$ significance levels, then combining inference from the two individual one sided tests for predictability would lead to an overall two sided test for predictability with asymptotic size of no greater than $\pi\%$.

6. Higher-order predictor serial correlation

We next consider how our procedures should be adapted to take account of possible additional serial correlation in the process for the predictor series x_t . To that end, we generalize the $AR(1)$ formulation placed on s_t in (2) to the $AR(p+1)$ formulation,

$$s_t = \phi s_{t-1} + v_t, \quad \psi(L)v_t = \epsilon_{xt} \quad (10)$$

where $\psi(L) := 1 + \sum_{j=1}^p \psi_j L^j$ is a finite-order stationary $AR(p)$ polynomial such that all of the roots of $\psi(z) = 0$ lie outside the unit circle, $|z| = 1$. The assumption of a finite-order autoregression for v_t , and hence s_t , appears to be standard in both the control variable and residual-augmented strands of the predictive regression literature; see, for example, [Campbell and Yogo \(2006\)](#), [Elliott \(2011\)](#), and [Demetrescu and Rodrigues \(2022\)](#), all of whom assume finite-order autoregressions. As argued in [Demetrescu and Rodrigues \(2022, p.431\)](#), in practice we might view this as an approximation to a more general linear process for v_t , although formally this would require establishing a suitable rate at which $p \rightarrow \infty$ as $T \rightarrow \infty$. We conjecture that the conventional rate conditions on p associated with unit root test statistics given in, for example, [Chang and Park \(2002\)](#), should suffice for this purpose.

To accommodate the additional stationary serial correlation introduced through (10), the following modifications to the procedures outlined previously need to be made. First, NB_{OLS} and NB need to be

based on the corresponding estimated *augmented* Dickey-Fuller regressions

$$\Delta x_t = \hat{\mu} + \hat{\phi}x_{t-1} + \sum_{i=1}^p \hat{\gamma}_i \Delta x_{t-i} + \hat{\epsilon}_{xt} \quad (11)$$

and

$$\Delta x_t = \tilde{\phi}(x_{t-1} - \tilde{\alpha}_x) + \sum_{i=1}^p \tilde{\gamma}_i \Delta x_{t-i} + \tilde{\epsilon}_{xt} \quad (12)$$

respectively, with the statistics re-defined as $NB_{OLS} := T\hat{\phi}/(1 - \sum_{i=1}^p \hat{\gamma}_i)$ and $NB := T\tilde{\phi}/(1 - \sum_{i=1}^p \tilde{\gamma}_i)$. Each of these ADF regressions now includes p lagged difference terms. In practice p can be chosen by any consistent lag selection method; the numerical simulations reported in Section 7 use the MBIC rule of Ng and Perron (2001). Next, the residual $\hat{\epsilon}_{xt}$ from (11) is used to calculate $\hat{\rho}_{xy}$ in (8), and finally, the residual $\tilde{\epsilon}_{xt}$ from (12) enters the regression (Eq. 7) for calculating \mathcal{J}^* . With these modifications implemented, the hybrid procedure outlined in Section 5 can continue to be implemented using the same set of conservative critical values.

7. Finite sample simulations

In this section, we evaluate the finite sample size and power properties of the \mathcal{J}_{hyb}^w procedure developed in Section 5. We generate data using a sample size $T = 200$ from the model (1)–(3) with $(e_{1t}, e_{2t})' \sim IIDN(0, I_2)$, $\sigma_x = \sigma_y = 1$ and drawing s_1 as a standard normal variate. We set $\alpha_y = \alpha_x = 0$ as the tests we calculate are invariant to these constant terms. The values of ρ_{xy} , c and g we consider are the same as in the asymptotic analysis of Figs. 1–4 to facilitate a comparison between finite sample and asymptotic performance. Upper tail 0.05-level tests are again conducted, with the results based on 20,000 replications. Throughout, we estimate p using the MBIC rule of Ng and Perron (2001) with a maximum permitted lag order of $p_{\max} = \lfloor 12(T/100)^{1/4} \rfloor$ ($\lfloor \cdot \rfloor$ denoting the integer part) together with the modification suggested by Perron and Qu (2007).

We first consider simulations of the finite sample size of \mathcal{J}_{hyb}^w , i.e., setting $g = 0$, allowing for additional serial correlation in the process for x_t through the specification

$$s_t = \phi s_{t-1} + v_t \quad (13)$$

$$v_t = \varphi v_{t-1} + e_{1t} - \theta e_{1,t-1} \quad (14)$$

for various values of $\varphi, \theta \in \{-0.5, 0, 0.5\}$. The simulation DGP for v_t in (14) allows for MA behavior whenever $\theta \neq 0$. We recall that this is not formally allowed for in the DGP specified for v_t in (10), but it is still of interest to consider MA errors in the simulations to investigate how well our proposed tests work in such cases, not least given our conjecture that the tests will remain valid for MA errors under a suitable rate condition on p .

Table 4 reports the results across the different settings for c and ρ_{xy} . We observe size to be generally well-controlled and close to the nominal level, with the serial correlation parameter settings for ϕ and θ having relatively little bearing on the rejection frequency under the null. Some modest over-size is apparent for the more negative values of ρ_{xy} when c is zero or small, but this diminishes as c increases and as the model innovations become less correlated.

That the over-size is most apparent around $c = 0$ and $\rho_{xy} = -0.95$ is a consequence of (i) in this region the \mathcal{J}_{hyb}^w procedure will be almost always be performing \mathcal{J}_{con}^w and (ii) as noted above $c\nu_{\pi}^w(\rho_{xy})$ is obtained with $c = 0$ (i.e., \mathcal{J}_{con}^w is asymptotically correctly sized for $c = 0$ and conservative elsewhere). Consequently, when finite sample over-size of \mathcal{J}_{hyb}^w occurs via \mathcal{J}_{con}^w it would be expected to be most prominent when $c = 0$. Of course, this does not resolve why finite sample over-size (as opposed to near correct- or under-size) should be manifest in the first place and we have no ready explanation for this. What we can say is that the over-size diminishes reasonably quickly as the sample size increases and the

Table 4. Finite sample size of nominal 0.05-level \mathcal{J}_{hyb}^w tests, $T = 200$.

ϕ	θ	$c = 0$	$c = 1.25$	$c = 2.5$	$c = 5$	$c = 10$	$c = 25$	$c = 50$	$c = 100$
<i>Panel A. $\rho_{xy} = -0.95$</i>									
0.0	0.0	0.075	0.067	0.065	0.061	0.052	0.038	0.056	0.057
0.0	0.5	0.072	0.063	0.058	0.050	0.040	0.035	0.047	0.053
0.0	-0.5	0.078	0.070	0.067	0.065	0.059	0.046	0.035	0.065
0.5	0.0	0.075	0.069	0.069	0.066	0.061	0.052	0.039	0.036
-0.5	0.0	0.076	0.066	0.062	0.055	0.045	0.031	0.049	0.050
0.5	-0.5	0.080	0.071	0.070	0.068	0.061	0.052	0.044	0.038
-0.5	0.5	0.077	0.061	0.051	0.041	0.032	0.030	0.043	0.050
<i>Panel B. $\rho_{xy} = -0.7$</i>									
0.0	0.0	0.068	0.064	0.063	0.059	0.054	0.043	0.050	0.054
0.0	0.5	0.064	0.060	0.057	0.051	0.043	0.036	0.046	0.049
0.0	-0.5	0.070	0.066	0.064	0.062	0.058	0.052	0.042	0.060
0.5	0.0	0.068	0.064	0.063	0.062	0.058	0.054	0.046	0.040
-0.5	0.0	0.068	0.064	0.061	0.056	0.048	0.035	0.045	0.048
0.5	-0.5	0.072	0.066	0.065	0.063	0.058	0.053	0.049	0.043
-0.5	0.5	0.062	0.058	0.052	0.044	0.034	0.032	0.042	0.049
<i>Panel C. $\rho_{xy} = -0.5$</i>									
0.0	0.0	0.061	0.060	0.059	0.057	0.055	0.047	0.048	0.052
0.0	0.5	0.059	0.058	0.056	0.051	0.046	0.038	0.045	0.049
0.0	-0.5	0.063	0.061	0.060	0.059	0.058	0.053	0.046	0.056
0.5	0.0	0.063	0.060	0.059	0.058	0.057	0.056	0.049	0.042
-0.5	0.0	0.062	0.060	0.059	0.056	0.050	0.039	0.044	0.048
0.5	-0.5	0.065	0.061	0.060	0.060	0.057	0.055	0.052	0.046
-0.5	0.5	0.060	0.057	0.054	0.046	0.039	0.034	0.043	0.049
<i>Panel D. $\rho_{xy} = -0.1$</i>									
0.0	0.0	0.053	0.049	0.049	0.050	0.049	0.046	0.045	0.048
0.0	0.5	0.051	0.049	0.050	0.050	0.048	0.044	0.045	0.048
0.0	-0.5	0.053	0.050	0.050	0.049	0.049	0.048	0.045	0.047
0.5	0.0	0.052	0.051	0.050	0.050	0.049	0.048	0.047	0.043
-0.5	0.0	0.051	0.050	0.050	0.050	0.050	0.044	0.044	0.047
0.5	-0.5	0.053	0.052	0.051	0.050	0.050	0.049	0.049	0.046
-0.5	0.5	0.052	0.051	0.049	0.049	0.046	0.042	0.047	0.050

asymptotics start to assert themselves. For example, the leading entry of Table 4 ($c = 0$, $\rho_{xy} = -0.95$) shows a size of 0.075 when $T = 200$. For $T = 400$ and $T = 800$ the corresponding sizes are 0.060 and 0.056, respectively.

We next evaluate the finite sample power of \mathcal{J}_{hyb}^w , and do so through comparison with the \mathcal{J}_{hyb} procedure of HLT and the test procedure proposed by EMW, which we denote by EMW . As discussed in Section 1, the EMW test procedure is the most natural extant comparator for the \mathcal{J}_{hyb}^w test, given that both exclude explosive predictors (recall that HLT's \mathcal{J}_{hyb} procedure allows for a small degree of local explosivity). Here we set $\varphi = \theta = 0$, but do not assume knowledge of this and continue to determine p using the method of the previous section. To implement the EMW procedure, we adopt the switching function specified on p.799 of EMW so that the standard t -test, \mathcal{J}_N , is applied if an estimate of the local offset c is at least 130, while their weighted average power criterion-based test is applied otherwise, using the sample statistics and long run correlation estimator specified on p.697 of Jansson and Moreira (2006). To estimate c we follow HLT and use $-T\hat{\phi}$ from (11), and when the standard t -test is used in EMW , we follow EMW's approach of setting the critical value to the usual value of 1.645 for non negative estimates of the long run correlation parameter, but to set it to 1.7 for negative estimates. Long run variances are calculated using a Bartlett kernel with lag truncation $\lfloor T^{1/3} \rfloor$. The \mathcal{J}_{hyb} procedure was implemented as in HLT, i.e., the same procedure as \mathcal{J}_{hyb}^w in Section 5 above, but with \mathcal{J}'_{con} replacing \mathcal{J}_{con}^w .

We first consider the comparison between \mathcal{J}_{hyb}^w and \mathcal{J}_{hyb} . Figure 5 gives the results for $\rho_{xy} = -0.95$. In Fig. 5(a) where $c = 0$, we see that \mathcal{J}_{hyb}^w and \mathcal{J}_{hyb} have approximately the same size, slightly above the

nominal level (cf. Table 4 for \mathcal{J}_{hyb}^w). It is clear however that \mathcal{J}_{hyb}^w is substantially more powerful than \mathcal{J}_{hyb} . The relevant asymptotic counterpart here is a comparison between \mathcal{J}_{con}^w and \mathcal{J}'_{con} in Fig. 1(a) and the power differences there appear even more significant, suggesting that we would see further power gains of \mathcal{J}_{hyb}^w over \mathcal{J}_{hyb} for finite sample sizes larger than $T = 200$. Elsewhere in Fig. 5 we see that \mathcal{J}_{hyb}^w continues to be more powerful than \mathcal{J}_{hyb} for values of c up to $c = 5$; thereafter \mathcal{J}_{hyb}^w and \mathcal{J}_{hyb} have identical power profiles. Notice again that the slight over-size associated with the tests is less apparent for the larger values of c . Figures 6–8 show the results for the other values of ρ_{xy} . For small c we continue to see \mathcal{J}_{hyb}^w outperform \mathcal{J}_{hyb} , while they behave similarly elsewhere, again in line with the asymptotic comparisons of \mathcal{J}_{con}^w and \mathcal{J}'_{con} in Figs. 2–4. Clearly then, there are potential benefits to be gained in practice by using the procedure \mathcal{J}_{hyb}^w instead of \mathcal{J}_{hyb} , since it performs either as well as or better than \mathcal{J}_{hyb}^w , offering power gains when the predictor variable is highly persistent, without worry of compromise when less persistent predictors are employed.

Finally, comparing our \mathcal{J}_{hyb}^w procedure to *EMW*, we find that, across the ρ_{xy} values we consider, the tests have very similar levels of power for the smaller values of c , while for larger c , \mathcal{J}_{hyb}^w clearly emerges as the more powerful procedure, with increasing gains over *EMW* seen as the magnitude of c increases. This feature is most apparent for the most negative values of ρ_{xy} , where the power gains of \mathcal{J}_{hyb}^w relative to *EMW* are apparent even for $c = 2.5$. Overall, in addition to offering a generally superior power profile to \mathcal{J}_{hyb} , the \mathcal{J}_{hyb}^w procedure can achieve substantial power advantages over the procedure of *EMW* for a wide range of ρ_{xy} and c combinations, while the reverse is never true.

8. Conclusions

In this article, we have proposed a new hybrid procedure designed to test for predictability in returns which is valid in cases where the predictor is either weakly or strongly persistent. Our proposed hybrid test is a complement to the closely related hybrid testing procedure of HLT. In particular, the simulation results presented in HLT highlight that their hybrid test outperforms other extant predictability tests in most settings, but is outperformed by the test procedure of *EMW* in the case of strongly persistent predictors with the persistence parameter c either zero or small. The comparison is, however, not on a level playing field because *EMW* rule out the possibility of mild explosivity ($c < 0$) in the predictor, while HLT allow for some mild explosivity. By restricting the predictor to be non explosive, we are able to consider using a control variable based test (in the spirit of Elliott, 2011), whereby the predictive regression is augmented by a GLS-based proxy for the innovation driving the predictor. We have shown that a feasible conservative implementation of this augmented test improves upon the asymptotic local power of the feasible test used in HLT, which is based on using a quasi-GLS demeaned version of the predictor (but no covariate), in precisely the region of the parameter space where the HLT procedure is less powerful than the *EMW* test. Moreover, we show that a test based on a weighted average of the augmented statistic and the quasi-GLS statistic from HLT delivers notable further improvements in asymptotic local power in this region. Our hybrid test then replaces the quasi-GLS test used in the hybrid procedure in HLT with this weighted test. Like the hybrid tests in both *EMW* and HLT our proposed hybrid test procedure reverts to a conventional regression t -test on the predictor (comparing to standard normal critical values) if the data suggest that the predictor is weakly persistent. Monte Carlo simulations presented in this article demonstrate that our proposed hybrid procedure is overall more powerful than both the *EMW* and HLT test procedures across a wide spectrum of values of the persistence level in the predictive regressor (including where c is zero or small) and the correlation coefficient between the innovations in the model. Where explosive predictors can be ruled out, we therefore recommend the procedure developed in this paper. Otherwise we recommend using the corresponding procedure we developed in HLT.

Acknowledgments

We dedicate this paper to the memory of Michael McAleer. Mike was a great inspiration to all three of us and he will be sadly missed. We thank an Associate Editor and two anonymous referees for their helpful and constructive comments on earlier versions of this paper.

Funding

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

References

- Breitung, J., Demetrescu, M. (2015). Instrumental variable and variable addition based inference in predictive regressions. *Journal of Econometrics* 187(1):358–375. doi:10.1016/j.jeconom.2013.10.018
- Campbell, J., Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics* 81(1):27–60. doi:10.1016/j.jfineco.2005.05.008
- Cavanagh, C. L., Elliott, G., Stock, J. H. (1995). Inference in models with nearly integrated regressors. *Econometric Theory* 11(5):1131–1147. doi:10.1017/S0266466600009981
- Chang, Y., Park, J. Y. (2002). On the asymptotics of ADF tests for unit roots. *Econometric Reviews* 21:431–447.
- Demetrescu, M., Rodrigues, P. M. M. (2022). Residual-augmented IVX predictive regression. *Journal of Econometrics* 227:429–460.
- Elliott, G. (2011). A control function approach for testing the usefulness of trending variables in forecast models and linear regression. *Journal of Econometrics* 164(1):79–91. doi:10.1016/j.jeconom.2011.02.014
- Elliott, G., Müller, U. K., Watson, M. W. (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica* 83(2):771–811. doi:10.3982/ECTA10535
- Elliott, G., Rothenberg, T. J., Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica* 64(4):813. doi:10.2307/2171846
- Fama, E. F. (1981). Stock returns, real activity, inflation, and money. *American Economic Review* 71:545–565.
- Goyal, A., Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management Science* 49(5):639–654. doi:10.1287/mnsc.49.5.639.15149
- Harvey, D. I., Leybourne, S. J., Taylor, A. R. (2021). Simple tests for stock return predictability with good size and power properties. *Journal of Econometrics* 224(1):198–214. doi:10.1016/j.jeconom.2021.01.004
- Jansson, M., Moreira, M. J. (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica* 74(3):681–714. doi:10.1111/j.1468-0262.2006.00679.x
- Kostakis, A., Magdalinos, T., Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies* 28(5):1506–1553. doi:10.1093/rfs/hhu139
- Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics* 74(2):209–235. doi:10.1016/j.jfineco.2002.11.002
- Nelson, C. R., Kim, M. J. (1993). Predictable stock returns: The role of small sample bias. *Journal of Finance* 48:641–661.
- Ng, S., Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69:1519–1554.
- Perron, P., Qu, Z. (2007). A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. *Economics Letters* 94:12–19.
- Phillips, P. C. B., Magdalinos, T. (2009). *Econometric inference in the vicinity of unity*. Working paper. Singapore: Management University.
- Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics* 54:375–421.
- Welch, I., Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21(4):1455–1508. doi:10.1093/rfs/hhm014

A. Appendix

Proof of Theorem 1. In what follows we may set $\alpha_y = \alpha_x = 0$ without loss of generality, by virtue of the exact invariance of \mathcal{J}^* to these parameters. We adopt the generic notation of (5) and (6), and denote OLS estimators of the parameter β and δ by $\hat{\beta}$ and $\hat{\delta}$, respectively, with sample means denoted by $\bar{y} = (T - 1)^{-1} \sum_{t=2}^T y_t$ and

$\bar{x}_{-1} = (T-1)^{-1} \sum_{t=2}^T x_{t-1}$. Summations are taken over $t = 2, \dots, T$, and integrals over $[0, 1]$, unless otherwise stated. We will make use of the following weak convergence results:

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} e_{1t} \Rightarrow W_1(r), \quad T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} e_{2t} \Rightarrow W_2(r)$$

where $[W_1(r), W_2(r)]'$ is a bivariate standard Brownian Motion process. Then we can also write

$$\begin{aligned} T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \epsilon_{xt} &\Rightarrow \sigma_x W_1(r) \\ T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \epsilon_{yt} &\Rightarrow \sigma_y \left\{ \rho_{xy} W_1(r) + \sqrt{1 - \rho_{xy}^2} W_2(r) \right\} \end{aligned}$$

and, under [Assumption S](#), $T^{-1/2} x_{\lfloor Tr \rfloor} \Rightarrow \sigma_x \int_0^r e^{-(r-s)c} dW_1(s)$.

We will make use of the following easily established results: $T^{-1} \sum (y_t - \bar{y})^2 \xrightarrow{p} \sigma_y^2$, $\tilde{\alpha}_x = x_1 + o_p(1)$, $T^{-1} \sum \tilde{\epsilon}_{xt}^2 \xrightarrow{p} \sigma_x^2$. Next,

$$\begin{aligned} T\tilde{\phi} &\Rightarrow \frac{\int W_{1c}(r) dW_{1c}(r)}{\int W_{1c}(r)^2 dr} \\ &= c + \frac{\int W_{1c}(r) dW_1(r)}{\int W_{1c}(r)^2 dr} := \mathcal{S}^{NB}(c) \end{aligned}$$

$$\begin{aligned} T^{-1} \sum (x_{t-1} - \bar{x}_{-1}) \tilde{\epsilon}_{xt} &= T^{-1} \sum (x_{t-1} - \bar{x}_{-1}) \{ \Delta x_t - \tilde{\phi}(x_{t-1} - x_1) \} + o_p(1) \\ &= T^{-1} \sum (x_{t-1} - \bar{x}_{-1}) \Delta x_t - T\tilde{\phi} T^{-2} \sum (x_{t-1} - \bar{x}_{-1}) x_{t-1} + o_p(1) \\ &\Rightarrow \sigma_x^2 \int \bar{W}_{1c}(r) dW_{1c}(r) - \sigma_x^2 \mathcal{S}^{NB}(c) \int \bar{W}_{1c}(r)^2 dr \end{aligned}$$

$$T^{-1} \sum (x_{t-1} - \bar{x}_{-1}) y_t \Rightarrow g\sigma_x \int \bar{W}_{1c}(r)^2 dr + \sigma_x \sigma_y \int \bar{W}_{1c}(r) d \left\{ \rho_{xy} W_1(r) + \sqrt{1 - \rho_{xy}^2} W_2(r) \right\}$$

$$\begin{aligned} T^{-1} \sum \tilde{\epsilon}_{xt} (y_t - \bar{y}) &= T^{-1} \sum \tilde{\epsilon}_{xt} (gT^{-1}(x_{t-1} - \bar{x}_{-1}) + \epsilon_{yt} - \bar{\epsilon}_y) \\ &= gT^{-2} \sum (x_{t-1} - \bar{x}_{-1}) \tilde{\epsilon}_{xt} + T^{-1} \sum \tilde{\epsilon}_{xt} (\epsilon_{yt} - \bar{\epsilon}_y) \\ &= T^{-1} \sum \{ \Delta(x_t - x_1) - \tilde{\phi}(x_{t-1} - x_1) \} (\epsilon_{yt} - \bar{\epsilon}_y) + o_p(1) \\ &= T^{-1} \sum (\epsilon_{yt} - \bar{\epsilon}_y) \Delta x_t + o_p(1) \Rightarrow \sigma_{xy} \end{aligned}$$

$$\begin{aligned} T^{-1} \sum \hat{v}_t^2 &= T^{-1} \sum \{ (y_t - \bar{y}) - \hat{\beta}(x_{t-1} - \bar{x}_{-1}) - \hat{\delta}(\tilde{\epsilon}_{xt} - \bar{\tilde{\epsilon}}_x) \}^2 \\ &= T^{-1} \sum (y_t - \bar{y})^2 + \hat{\beta}^2 T^{-1} \sum (x_{t-1} - \bar{x}_{-1})^2 + \hat{\delta}^2 T^{-1} \sum \tilde{\epsilon}_{xt}^2 \\ &\quad - 2\hat{\beta} T^{-1} \sum (y_t - \bar{y})(x_{t-1} - \bar{x}_{-1}) - 2\hat{\delta} T^{-1} \sum (y_t - \bar{y}) \tilde{\epsilon}_{xt} + 2\hat{\beta}\hat{\delta} T^{-1} \sum (x_{t-1} - \bar{x}_{-1}) \tilde{\epsilon}_{xt} \\ &\quad + o_p(1) \\ &= T^{-1} \sum (y_t - \bar{y})^2 + \hat{\delta}^2 T^{-1} \sum \tilde{\epsilon}_{xt}^2 - 2\hat{\delta} T^{-1} \sum (y_t - \bar{y}) \tilde{\epsilon}_{xt} + o_p(1) \\ &\Rightarrow \sigma_y^2 + (\sigma_{xy}/\sigma_x^2)^2 \sigma_x^2 - 2(\sigma_{xy}/\sigma_x^2) \sigma_{xy} = \sigma_y^2 (1 - \rho_{xy}^2) \end{aligned}$$

using the fact that $\bar{\tilde{\epsilon}}_x = T^{-1} \sum \tilde{\epsilon}_{xt} = T^{-1} \sum \Delta(x_t - x_1) - T\tilde{\phi}T^{-2} \sum(x_{t-1} - x_1) + o_p(1) \Rightarrow 0$. Then,

$$\begin{aligned}
\begin{bmatrix} \hat{\beta} \\ \hat{\delta} \end{bmatrix} &= \begin{bmatrix} \sum(x_{t-1} - \bar{x}_{-1})^2 & \sum(x_{t-1} - \bar{x}_{-1})\tilde{\epsilon}_{xt} \\ \sum(x_{t-1} - \bar{x}_{-1})\tilde{\epsilon}_{xt} & \sum\tilde{\epsilon}_{xt}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum(x_{t-1} - \bar{x}_{-1})y_t \\ \sum\tilde{\epsilon}_{xt}(y_t - \bar{y}) \end{bmatrix} \\
\begin{bmatrix} T\hat{\beta} \\ \hat{\delta} \end{bmatrix} &= \left(\begin{bmatrix} T^{-1} & 0 \\ 0 & T^{-1} \end{bmatrix} \begin{bmatrix} \sum(x_{t-1} - \bar{x}_{-1})^2 & \sum(x_{t-1} - \bar{x}_{-1})\tilde{\epsilon}_{xt} \\ \sum(x_{t-1} - \bar{x}_{-1})\tilde{\epsilon}_{xt} & \sum\tilde{\epsilon}_{xt}^2 \end{bmatrix} \begin{bmatrix} T^{-1} & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \\
&\quad \cdot \begin{bmatrix} T^{-1} \sum(x_{t-1} - \bar{x}_{-1})y_t \\ T^{-1} \sum\tilde{\epsilon}_{xt}(y_t - \bar{y}) \end{bmatrix} \\
&= \begin{bmatrix} T^{-2} \sum(x_{t-1} - \bar{x}_{-1})^2 & T^{-1} \sum(x_{t-1} - \bar{x}_{-1})\tilde{\epsilon}_{xt} \\ T^{-2} \sum(x_{t-1} - \bar{x}_{-1})\tilde{\epsilon}_{xt} & T^{-1} \sum\tilde{\epsilon}_{xt}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-1} \sum(x_{t-1} - \bar{x}_{-1})y_t \\ T^{-1} \sum\tilde{\epsilon}_{xt}(y_t - \bar{y}) \end{bmatrix} \\
&\Rightarrow \begin{bmatrix} \sigma_x^2 \int \bar{W}_{1c}(r)^2 dr & \sigma_x^2 \int \bar{W}_{1c}(r) dW_{1c}(r) - \sigma_x^2 S^{NB}(c) \int \bar{W}_{1c}(r)^2 dr \\ 0 & \sigma_x^2 \end{bmatrix}^{-1} \quad (A.1) \\
&\quad \cdot \begin{bmatrix} g\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr + \sigma_x\sigma_y \int \bar{W}_{1c}(r) d \left\{ \rho_{xy}W_1(r) + \sqrt{1 - \rho_{xy}^2}W_2(r) \right\} \\ \sigma_{xy} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr} & -\frac{\int \bar{W}_{1c}(r) dW_{1c}(r) - S^{NB}(c) \int \bar{W}_{1c}(r)^2 dr}{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr} \\ 0 & 1/\sigma_x^2 \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} g\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr + \sigma_x\sigma_y \int \bar{W}_{1c}(r) d \left\{ \rho_{xy}W_1(r) + \sqrt{1 - \rho_{xy}^2}W_2(r) \right\} \\ \sigma_{xy} \end{bmatrix} \\
&= \begin{bmatrix} \frac{g\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr + \sigma_x\sigma_y \int \bar{W}_{1c}(r) d \left\{ \rho_{xy}W_1(r) + \sqrt{1 - \rho_{xy}^2}W_2(r) \right\}}{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr} - \sigma_{xy} \frac{\int \bar{W}_{1c}(r) dW_{1c}(r) - S^{NB}(c) \int \bar{W}_{1c}(r)^2 dr}{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr} \\ \sigma_{xy}/\sigma_x^2 \end{bmatrix}.
\end{aligned}$$

The expression for $T\hat{\beta}$ can be simplified as follows:

$$\begin{aligned}
T\hat{\beta} &\Rightarrow g + \frac{\sigma_x\sigma_y \int \bar{W}_{1c}(r) d \left\{ \rho_{xy}W_1(r) + \sqrt{1 - \rho_{xy}^2}W_2(r) \right\}}{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr} - \sigma_{xy} \frac{\int \bar{W}_{1c}(r) dW_{1c}(r) - S^{NB}(c) \int \bar{W}_{1c}(r)^2 dr}{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr} \\
&= g + \frac{\sigma_x\sigma_y \int \bar{W}_{1c}(r) d \left\{ \rho_{xy}W_1(r) + \sqrt{1 - \rho_{xy}^2}W_2(r) \right\}}{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr} - \sigma_{xy} \left(\frac{\int \bar{W}_{1c}(r) dW_{1c}(r)}{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr} - \frac{S^{NB}(c)}{\sigma_x^2} \right) \\
&= g + \frac{\sigma_x\sigma_y \sqrt{1 - \rho_{xy}^2} \int \bar{W}_{1c}(r) dW_2(r)}{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr} + \frac{\sigma_x\sigma_y \rho_{xy} \int \bar{W}_{1c}(r) dW_1(r)}{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr} \\
&\quad - \frac{\sigma_{xy}}{\sigma_x^2} \left(c + \frac{\int \bar{W}_{1c}(r) dW_1(r)}{\int \bar{W}_{1c}(r)^2 dr} - S^{NB}(c) \right) \\
&= g + \frac{\sigma_x\sigma_y \sqrt{1 - \rho_{xy}^2} \int \bar{W}_{1c}(r) dW_2(r)}{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr} - \frac{\sigma_{xy}}{\sigma_x^2} \left(c - c - \frac{\int W_{1c}(r) dW_1(r)}{\int W_{1c}(r)^2 dr} \right) \\
&= g + \frac{\sigma_x\sigma_y \sqrt{1 - \rho_{xy}^2} \int \bar{W}_{1c}(r) dW_2(r)}{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr} + \frac{\sigma_{xy} \int W_{1c}(r) dW_1(r)}{\sigma_x^2 \int W_{1c}(r)^2 dr}.
\end{aligned}$$

Hence, using m to denote element [1,1] of the inverse matrix in (A.1), we find

$$\begin{aligned}
 \mathcal{J}^* &= \frac{T\hat{\beta}}{\sqrt{T^{-1} \sum \hat{v}_t^2 T^2 m}} \\
 &\Rightarrow \frac{g + \frac{\sigma_x \sigma_y \sqrt{1 - \rho_{xy}^2} \int \bar{W}_{1c}(r) dW_2(r)}{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr} + \frac{\sigma_{xy} \int W_{1c}(r) dW_1(r)}{\sigma_x^2 \int W_{1c}(r)^2 dr}}{\sqrt{\frac{\sigma_y^2 (1 - \rho_{xy}^2)}{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr}}} \\
 &= \left[g + \frac{\sigma_x \sigma_y \sqrt{1 - \rho_{xy}^2} \int \bar{W}_{1c}(r) dW_2(r)}{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr} + \frac{\sigma_{xy} \int W_{1c}(r) dW_1(r)}{\sigma_x^2 \int W_{1c}(r)^2 dr} \right] \sqrt{\frac{\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr}{\sigma_y^2 (1 - \rho_{xy}^2)}} \\
 &= \frac{g \sigma_x \sqrt{\int \bar{W}_{1c}(r)^2 dr}}{\sigma_y \sqrt{1 - \rho_{xy}^2}} + \frac{\int \bar{W}_{1c}(r) dW_2(r)}{\sqrt{\int \bar{W}_{1c}(r)^2 dr}} + \frac{\rho_{xy} \sqrt{\int \bar{W}_{1c}(r)^2 dr} \int W_{1c}(r) dW_1(r)}{\sqrt{1 - \rho_{xy}^2} \int W_{1c}(r)^2 dr}.
 \end{aligned}$$

□