

Bonferroni Type Tests for Return Predictability and the Initial Condition

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

To cite this article: Sam Astill, David I. Harvey, Stephen J. Leybourne & A. M. Robert Taylor (2023): Bonferroni Type Tests for Return Predictability and the Initial Condition, Journal of Business & Economic Statistics, DOI: [10.1080/07350015.2023.2201313](https://doi.org/10.1080/07350015.2023.2201313)

To link to this article: <https://doi.org/10.1080/07350015.2023.2201313>



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



[View supplementary material](#)



Published online: 30 May 2023.



[Submit your article to this journal](#)



Article views: 129



[View related articles](#)



[View Crossmark data](#)

Bonferroni Type Tests for Return Predictability and the Initial Condition

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

^aEssex Business School, University of Essex, Colchester, UK; ^bGranger Centre for Time Series Econometrics and School of Economics, University of Nottingham, Nottingham, UK

ABSTRACT

We develop tests for predictability that are robust to both the magnitude of the initial condition and the degree of persistence of the predictor. While the popular Bonferroni Q test of Campbell and Yogo displays excellent power properties for strongly persistent predictors with an asymptotically negligible initial condition, it can suffer from severe size distortions and power losses when either the initial condition is asymptotically non-negligible or the predictor is weakly persistent. The Bonferroni t test of Cavanagh, Elliott, and Stock, although displaying power well below that of the Bonferroni Q test for strongly persistent predictors with an asymptotically negligible initial condition, displays superior size control and power when the initial condition is asymptotically nonnegligible. In the case where the predictor is weakly persistent, a conventional regression t test comparing to standard normal quantiles is known to be asymptotically optimal under Gaussianity. Based on these properties, we propose two asymptotically size controlled hybrid tests that are functions of the Bonferroni Q , Bonferroni t , and conventional t tests. Our proposed hybrid tests exhibit very good power regardless of the magnitude of the initial condition or the persistence degree of the predictor. An empirical application to the data originally analyzed by Campbell and Yogo shows our new hybrid tests are much more likely to find evidence of predictability than the Bonferroni Q test when the initial condition of the predictor is estimated to be large in magnitude.

ARTICLE HISTORY

Received February 2022
Accepted April 2023

KEYWORDS

Bonferroni tests; Hybrid tests; Initial condition; Predictive regression; Unknown regressor persistence

1. Introduction and Motivation

Testing for the predictability of asset returns has been the subject of numerous studies in the applied economics and finance literature, assessing the predictive strength of a range of candidate predictor variables, including valuation ratios, interest rates and other financial and macroeconomic variables. Fama (1981) examines the predictability of stock returns using various candidate predictors including interest rates, industrial production, GNP and capital stock and expenditure, while Campbell and Yogo (2006) [CY, hereafter] consider candidate predictors that include the dividend and earnings price ratios, the three-month T-bill rate and the long-short yield spread. The standard approaches to determining whether returns are predictable are based on a simple linear predictive regression model with a constant and lagged putative predictor, which we denote as x_{t-1} , with slope coefficient β .

A common finding in empirical studies into return predictability is that the putative predictor is often both strongly persistent and endogenous, with a nonzero (often strongly negative) correlation between the errors in the predictive regression and the innovations driving the predictor process; see, *inter alia*, CY and Welch and Goyal (2008). In this situation Cavanagh, Elliott, and Stock (1995) [CES, hereafter] show that the standard t test on the estimate of β suffers from severe size distortions that are a function of both the degree of persistence and the endogeneity of the predictor. This finding has motivated

the development of numerous tests for predictability that are designed to allow for both strong persistence in the predictor series x_t , modeled by a first order autoregression with a local-to-unity coefficient $\rho = 1 - c/T$ (where c is an unknown finite constant and T is the sample size), and also predictor endogeneity. Arguably the most commonly employed test of this type is the Q test proposed by CY and it is this test that we will concentrate on in this article.¹

In brief, the Bonferroni Q test procedure of CY is based around computing a confidence interval for β using what is essentially a t -statistic obtained from the predictive regression augmented by the covariate $(x_t - \rho x_{t-1})$. When x_t is (near)-integrated, the local offset c in ρ is not consistently estimable, rendering the confidence interval calculation infeasible in practice. To overcome this problem, CY use a Bonferroni procedure, originally proposed in CES, whereby a confidence interval for ρ is first constructed by inverting the quasi-GLS demeaned Dickey-Fuller (ADF-GLS) unit root test of Elliott, Rothenberg,

¹Another strand of the literature focuses on instrumental variable estimation using an instrument which is less persistent than a local-to-unity process; see, *inter alia*, Kostakis, Magdalinos, and Stamatogiannis (2015) and Breitung and Demetrescu (2015). These tests are valid regardless of whether the predictor is weakly or strongly persistent, but are less powerful than the Q test when the predictor is strongly persistent, a significant drawback given a large number of the candidate predictors used in empirical work appear strongly persistent.

and Stock (1996) applied to the predictor, x_t . The bounds associated with this confidence interval for ρ are then used to deliver a feasible confidence interval for β .

For strongly persistent predictors, CY show that the Bonferroni Q test procedure has well controlled size and good power properties regardless of the value of the noncentrality parameter c and the degree of endogeneity of the predictor. These excellent empirical properties are, however, predicated on a key assumption that the initial condition of the predictor series, defined as the deviation of the initial value of the series from its underlying mean, is asymptotically negligible. In the context of stationary but near-integrated predictors, where the data are of order $O_p(T^{1/2})$, this assumption is only tenable if the beginning of the sample coincides with the beginning of the process. In practice this assumption is likely to be implausible; the predictors commonly considered will have been running for quite some time prior to the start of the observed sample. Consequently, it seems more plausible to allow the initial condition to be asymptotically nonnegligible (relative to the rest of the data on the predictor) and, as we will show, where this is the case it will influence the large sample properties of the Bonferroni Q test procedure. Exploring this issue forms the main focus of our article.

The Bonferroni Q procedure of CY relies on use of the ADF-GLS statistic to construct a confidence interval for ρ . Müller and Elliott (2003) show that the power of the ADF-GLS test against stationary alternatives is highly sensitive to the value of the initial condition. When the initial condition is of $o_p(T^{1/2})$, and hence asymptotically negligible, Müller and Elliott (2003) demonstrate that the ADF-GLS test has excellent power properties when ρ is near-integrated. However, where the initial condition is asymptotically nonnegligible, they show that the local alternative distribution of the ADF-GLS statistic is shifted to the right, relative to the asymptotically negligible initial condition case, leading to a reduction in relative power against left tailed alternatives, with this reduction more pronounced the larger is the absolute value of the initial condition. In the current predictive regression context, this leads to a rightwards shift in the confidence interval for ρ , and subsequently, a leftwards shift

in the confidence interval for β . Performing a right tailed test for predictability using the Bonferroni Q test, when a negative correlation exists between the innovations to returns and the predictor, entails examining whether the lower bound of the confidence interval for β exceeds zero, and so a leftward shift in this confidence interval induced by a large initial condition would be expected to result in a Q test that is undersized and lacking in power. Likewise, when performing a left tailed Q test, large values of the initial condition are anticipated to lead to oversizing in the Bonferroni Q test.

To illustrate, we now report results of a brief motivating empirical application to demonstrate the impact that initial conditions of different magnitudes can have on the Bonferroni Q test. Specifically, we examine 5%-level right tailed tests for predictability of the returns of the NYSE/AMEX value-weighted index from the Center for Research in Security Prices (CRSP) using the earnings-price ratio as a predictor, for the same monthly data from 1926M12 to 1994M12 as used in CY ($T = 817$). Based on the full sample of data, CY find that the earnings-price ratio is a significant predictor of returns. We repeat this exercise, but instead perform the Q test on data from $t = t_s, \dots, T$ across multiple start dates $t_s = 1926M12, \dots, 1945M12$. The results of this exercise are summarized in Figure 1. The red and green highlighted line plots, for each start date t_s , the lower bound of the confidence interval for β calculated from the right tailed Bonferroni Q test performed at the 5% nominal (asymptotic) level, with a lower bound above zero signaling a rejection (green highlights) and a lower bound below zero signaling nonrejection (red highlights). The blue line plots an estimate of the magnitude of the initial condition of the predictor variable relative to the variance for each subsample, $|\hat{\theta}|$ (subsequently defined in (15)), using the method proposed by Harvey and Leybourne (2005), and the grey shaded regions further highlight those start dates t_s for which the Bonferroni Q test fails to reject the null of no predictability.

It is apparent that while the null hypothesis of $\beta = 0$ is rejected by the Bonferroni Q test for the full sample (as indicated by the green highlighted line at $t_s=1926M12$), and for a majority

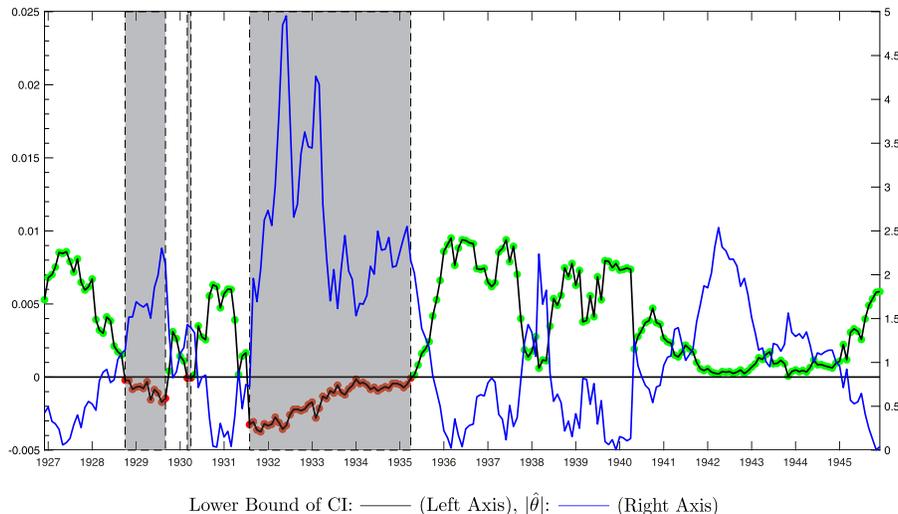


Figure 1. Lower bound of confidence interval of Bonferroni Q test and $|\hat{\theta}|$.

of other subsamples considered, there are a substantial number of start dates for which the Q test fails to find predictability. In general, it can be seen that for subsamples where the estimate of the relative magnitude of the initial condition is small in absolute value, the Q test rejects the null of no predictability, whereas in subsamples where this estimate is large in absolute value the Q test often fails to reject the null. These findings are in line with our conjecture that large initial conditions in the predictor will cause right tailed predictability tests to exhibit lower rejection frequencies. This is an important finding and suggests that the magnitude of the initial condition of the predictor is indeed an important consideration when applying the Bonferroni Q test to empirical data. Failing to account for the impact of the initial condition on this testing strategy can lead to different conclusions based on the particular subsample of data chosen, which is clearly an undesirable feature.

Motivated by these empirical findings, our aim in this article is to develop tests for predictability that offer a far greater degree of robustness to the initial condition of the predictor series than extant tests. It is important to stress that no CY/CES-type test can be invariant, even asymptotically, to an asymptotically nonnegligible initial condition (unless $c = 0$), because the magnitude of the initial condition features in the limiting distribution of both the unit root and predictive regression statistics used in the construction of the Bonferroni confidence intervals for ρ and β . The tests we develop are constructed so that their asymptotic size is controlled across the predictor's initial condition magnitude, degree of persistence and endogeneity, while also retaining most of the excellent power properties afforded by the Bonferroni Q test in the case where the initial condition is asymptotically negligible. Specifically, we propose hybrid test procedures based on the Bonferroni Q test of CY, the Bonferroni t test of CES, and a conventional predictive regression t test. While CY show that the Bonferroni t test displays poor power properties relative to the Bonferroni Q test when the predictor is driven by a local-to-unity process with an asymptotically negligible initial condition, we show that the size and power of the Bonferroni t test has the attractive feature of being relatively unaffected by whether the initial condition is asymptotically negligible or nonnegligible. This is, in part, because the Bonferroni t test of CES bases its confidence interval for ρ on the OLS demeaned Dickey-Fuller statistic (ADF-OLS), rather than the ADF-GLS statistic, and it is known from Müller and Elliott (2003) that ADF-OLS is considerably more robust than ADF-GLS to the initial condition.

We propose two approaches designed for strongly persistent predictors. The first is based on a union-of-rejections strategy in which the null of no predictability is rejected if either the Bonferroni Q test or Bonferroni t test rejects. Such a strategy is common in the time series econometrics literature, following Harvey, Leybourne, and Taylor (2009) in the context of unit root testing. The second uses the estimate of the magnitude of the initial condition relative to the variance proposed in Harvey and Leybourne (2005) to construct a weighted average of the Bonferroni Q and t tests, calibrated such that greater weight is placed on the Bonferroni Q test (t test) when the estimated magnitude of the initial condition is small (large). We will show that both of these combined tests are able to control asymptotic

size in the local-to-unity environment, regardless of the value of the initial condition, maintain power close to that of the Bonferroni Q test when the initial condition is small, and achieve power close to that of the Bonferroni t test when the magnitude of the initial condition is large.

While our primary analysis concerns a strongly persistent predictor (as in CY), it is also important to note that both the Bonferroni Q test of CY and Bonferroni t test of CES are (asymptotically) invalid if the predictor is weakly persistent ($|\rho| < 1$), with the confidence interval provided by inverting the ADF-GLS and ADF-OLS tests having zero asymptotic coverage for weakly persistent series. To ensure that our proposed test procedures are also robust to the possibility of weak persistence in the predictor, we adopt a hybrid testing approach, based on a similar switching strategy to those developed by Elliott, Müller, and Watson (2015) [EMW, hereafter] and Harvey, Leybourne, and Taylor (2021), whereby a conventional regression t test with standard normal critical values is implemented, rather than one of the combined tests outlined above, when there is sufficiently strong evidence to suggest that the predictor is weakly persistent, this test being asymptotically optimal (among feasible tests) under Gaussianity when the predictor is weakly dependent; see Jansson and Moreira (2006, p. 704).

The article is organized as follows. The predictive regression model and assumptions are detailed in Section 2. In Section 3, the Bonferroni Q test of CY and Bonferroni t test of CES are introduced and the asymptotic behavior of these tests is examined when the initial condition is asymptotically nonnegligible. Our proposed hybrid test procedures are outlined in Section 4, and their local asymptotic power is compared with those of the CY and CES tests. Section 5 reports results of an empirical application of our proposed hybrid test procedures to the dataset considered in CY. An on-line supplementary appendix provides additional local asymptotic power simulations, finite sample size and power simulations, additional empirical results, and a proof of our main technical result.

2. The Predictive Regression Model and Assumptions

We consider the following predictive regression model

$$r_t = \alpha + \beta x_{t-1} + u_t, \quad t = 1, \dots, T \quad (1)$$

where r_t denotes the (excess) return in period t , and x_{t-1} denotes a putative predictor observed at time $t - 1$. We assume the process for x_t is given by

$$x_t = \mu + w_t, \quad t = 0, \dots, T \quad (2)$$

$$w_t = \rho w_{t-1} + v_t, \quad t = 1, \dots, T \quad (3)$$

and make the following assumptions concerning the shocks u_t and v_t .

Assumption 1. We assume that $\psi(L)v_t = e_t$ where $\psi(L) := \sum_{i=0}^{p-1} \psi_i L^i$ with $\psi_0 = 1$ and $\psi(1) \neq 0$, with the roots of $\psi(L)$ assumed to be less than one in absolute value. We assume that $z_t := (u_t, e_t)'$ is a bivariate martingale difference sequence with respect to the natural filtration $\mathcal{F}_t := \sigma \{z_s, s \leq t\}$ satisfying the

following conditions: (i) $E[z_t z_t'] = \begin{bmatrix} \sigma_u^2 & \sigma_{ue} \\ \sigma_{ue} & \sigma_e^2 \end{bmatrix}$, (ii) $\sup_t E[u_t^4] < \infty$, and (iii) $\sup_t E[e_t^4] < \infty$. For future reference, we define $\omega_v^2 := \lim_{T \rightarrow \infty} T^{-1} E(\sum_{t=1}^T v_t)^2 = \sigma_e^2 / \psi(1)^2$ to be the long run variance of the error process $\{v_t\}$, and $\delta := \sigma_{ue} / \sigma_u \sigma_e$ as the correlation between the innovations $\{u_t\}$ and $\{e_t\}$.

Remark 2.1. The conditions in [Assumption 1](#) coincide with the most general set of assumptions considered in CY (see pp. 56–57 of CY). The assumptions placed on z_t allow the sequence of innovations to be conditionally heteroscedastic but imposes unconditional homoscedasticity. Notice that the MDS aspect of [Assumption 1](#) implies the standard assumption made in this literature that the unpredictable component of returns, u_t , is serially uncorrelated. [Assumption 1](#) allows the dynamics of the predictor variable to be captured by an $AR(p)$, with the degree of persistence of the predictor (strong or weak) controlled by the parameter ρ in (3), as will be formalized in [Assumptions S.1, S.2, S.3](#), and [W](#).

As discussed in [Section 1](#), our focus is on tests of the null hypothesis that $(r_t - \alpha)$ is a MDS and, hence, that r_t is not predictable by x_{t-1} ; that is, $H_0 : \beta = \beta_0 = 0$ in (1).² Our aim is to develop tests that offer reliable size and good power under different assumptions regarding the degree of persistence in the predictor variable x_t , and under different assumptions regarding the order of magnitude of the *initial condition* of x_t , given by $w_0 = x_0 - \mu$. We therefore allow the predictor process $\{x_t\}$ in (2) to satisfy one of the following four assumptions.

Assumption S.1. The predictor $\{x_t\}$ is strongly persistent, with the autoregressive parameter ρ in (3) given by $\rho = 1 - c/T$ with $c = 0$. The initial condition w_0 is unrestricted.

Assumption S.2. The predictor $\{x_t\}$ is strongly persistent, with ρ in (3) given by $\rho = 1 - c/T$ with c a finite nonzero constant. The initial condition is given by $w_0 = o_p(T^{1/2})$.

Assumption S.3. The predictor $\{x_t\}$ is strongly persistent, with ρ in (3) given by $\rho = 1 - c/T$ with c a finite positive constant. The initial condition is given by $w_0 = \theta \sigma_w$ where σ_w^2 denotes the short run variance of the process $\{w_t\}$ and $\theta \sim N(\mu_\theta I(\sigma_\theta^2 = 0), \sigma_\theta^2)$, where $I(\cdot)$ denotes the indicator function that takes a value of 1 when its argument is true, 0 otherwise. When $\sigma_\theta^2 > 0$ we further assume that the random variable θ is independent of z_t for all t .

Assumption W. The predictor $\{x_t\}$ is weakly persistent. The parameter ρ in (3) is fixed and bounded away from unity, $|\rho| < 1$. The initial condition is given by $w_0 = O_p(1)$.

Remark 2.2. Under [Assumption S.1](#), x_t is a pure unit root (or $I(1)$) process. No restrictions need to be placed on the initial condition here because all of the testing procedures discussed in this article are exact invariant to w_0 in the pure unit root case. Under [Assumptions S.2](#) and [S.3](#), x_t is specified to follow a

(strongly persistent) local-to-unity process, with the degree of persistence of the process controlled by c . For $c > 0$, x_t is a stationary but near-integrated process, while for $c < 0$, x_t is a (locally) explosive process. [Assumption S.2](#) specifies the initial condition of x_t to be asymptotically negligible, while [Assumption S.3](#), in the context of stationary near-integrated predictors ($c > 0$), sets the initial condition of x_t to be proportional to the standard deviation of the stationary process $\{w_t\}$ (as in Müller and Elliott 2003); this implies $\sigma_w^2 = \omega_v^2 T / 2c + o(T)$ and hence that w_0 is of $O_p(T^{1/2})$, that is, the initial condition is asymptotically nonnegligible. Here, θ controls the magnitude of the initial condition (relative to σ_w). If $\sigma_\theta^2 = 0$ then the initial condition is fixed and is given by $w_0 = \mu_\theta \sigma_w$. On the other hand, if $\sigma_\theta^2 > 0$ then the initial condition is random with $w_0 \sim N(0, \sigma_\theta^2 \sigma_w^2)$.

Remark 2.3. Under [Assumption S.2](#), we allow for the possibility of explosive predictors, $c < 0$, as in CY. However, it is important to initialize an explosive predictor at an asymptotically negligible initial value, because otherwise the behavior of the predictor becomes dominated by the initialization (increasingly so over time), something which is unlikely to be credible for macroeconomic and financial variables. Hence, we do not consider asymptotically nonnegligible initial conditions in the explosive case.

Remark 2.4. Under [Assumption W](#), x_t follows a stationary process and the initial condition is, correspondingly, assumed to be of $O_p(1)$, as would arise if, for example, the initial condition was proportional to σ_w^2 in the case $|\rho| < 1$.

In practice, when the putative predictor is near-integrated with $c > 0$ it is difficult to know which of [Assumptions S.2](#) and [S.3](#) is the more appropriate, as the initial condition w_0 is unobserved (as distinct from the initial *observation* x_0) and we would not know, a priori, whether the initial condition is “large” or “small.” As discussed in [Section 1](#), an argument could be made for [Assumption S.3](#) to hold in the strongly persistent case on the basis that the initial condition derives from the first *sample* observation on the predictor and so should be specified to have the same stochastic order as the rest of the sample data. As we will show, the local asymptotic powers of predictability tests depend on which of [Assumptions S.2](#) and [S.3](#) holds, and, under [Assumption S.3](#), on the magnitude of the initial condition. Hence, it is important to consider the behavior of predictive regression tests under different initial condition assumptions and magnitudes with near-integrated predictors.

3. Behavior of Bonferroni Q and t Tests under Assumption S.3

3.1. Bonferroni Q and t Tests

CY and CES propose testing for predictability based on Bonferroni procedures that make use of confidence intervals for the unknown autoregressive parameter $\rho = 1 - c/T$, with these confidence intervals constructed by inverting unit root tests.

²The methods which we outline in this article could equally well be used to test the generic null hypothesis that $\beta = \beta_0$ in (1), but as the focus in equity forecasting is on testing the null hypothesis of a zero coefficient on the lagged predictor we will restrict our attention to $\beta_0 = 0$.

Table 1. Parameters to deliver one-sided tests with maximum 5% asymptotic size.

δ	Q^{GLS}		t^{OLS}		U	W_1	W_2	δ	Q^{GLS}		t^{OLS}		U	W_1	W_2
	$\underline{\alpha}_1^Q$	$\bar{\alpha}_1^Q$	$\underline{\alpha}_1^t$	$\bar{\alpha}_1^t$	ξ	ξ	ξ		$\underline{\alpha}_1^Q$	$\bar{\alpha}_1^Q$	$\underline{\alpha}_1^t$	$\bar{\alpha}_1^t$	ξ	ξ	ξ
-0.999	0.050	0.055	0.020	0.035	0.28	0.85	0.55	-0.500	0.080	0.280	0.060	0.035	0.28	0.60	0.45
-0.975	0.055	0.080	0.025	0.035	0.28	0.85	0.56	-0.475	0.085	0.285	0.050	0.035	0.20	0.58	0.36
-0.950	0.055	0.100	0.025	0.040	0.38	0.75	0.50	-0.450	0.085	0.295	0.055	0.040	0.21	0.56	0.36
-0.925	0.055	0.115	0.025	0.040	0.34	0.73	0.50	-0.425	0.090	0.310	0.035	0.040	0.21	0.56	0.36
-0.900	0.060	0.130	0.025	0.035	0.34	0.71	0.46	-0.400	0.090	0.320	0.060	0.040	0.21	0.53	0.36
-0.875	0.060	0.140	0.025	0.035	0.32	0.72	0.45	-0.375	0.095	0.330	0.040	0.040	0.18	0.45	0.36
-0.850	0.060	0.150	0.025	0.035	0.30	0.71	0.43	-0.350	0.100	0.345	0.030	0.040	0.18	0.45	0.36
-0.825	0.060	0.160	0.025	0.035	0.35	0.71	0.46	-0.325	0.100	0.355	0.015	0.045	0.18	0.43	0.32
-0.800	0.065	0.170	0.025	0.035	0.33	0.71	0.46	-0.300	0.105	0.360	0.010	0.050	0.18	0.41	0.26
-0.775	0.065	0.180	0.030	0.035	0.36	0.72	0.56	-0.275	0.110	0.370	0.005	0.040	0.18	0.41	0.25
-0.750	0.065	0.190	0.025	0.035	0.36	0.72	0.56	-0.250	0.115	0.375	0.005	0.035	0.18	0.41	0.25
-0.725	0.065	0.195	0.025	0.035	0.36	0.71	0.57	-0.225	0.125	0.380	0.005	0.025	0.18	0.35	0.23
-0.700	0.070	0.205	0.025	0.035	0.37	0.71	0.57	-0.200	0.130	0.390	0.005	0.025	0.10	0.33	0.21
-0.675	0.070	0.215	0.025	0.035	0.34	0.71	0.57	-0.175	0.140	0.395	0.005	0.010	0.10	0.33	0.21
-0.650	0.070	0.225	0.025	0.035	0.34	0.67	0.55	-0.150	0.150	0.400	0.005	0.010	0.10	0.33	0.21
-0.625	0.075	0.230	0.025	0.035	0.34	0.67	0.51	-0.125	0.160	0.405	0.005	0.010	0.10	0.33	0.21
-0.600	0.075	0.240	0.030	0.035	0.34	0.65	0.51	-0.100	0.175	0.415	0.005	0.005	0.10	0.33	0.21
-0.575	0.075	0.250	0.035	0.035	0.32	0.65	0.45	-0.075	0.190	0.420	0.005	0.005	0.10	0.33	0.21
-0.550	0.080	0.260	0.035	0.035	0.30	0.65	0.45	-0.050	0.215	0.425	0.005	0.005	0.10	0.33	0.21
-0.525	0.080	0.270	0.045	0.035	0.28	0.65	0.45	-0.025	0.250	0.435	0.005	0.005	0.10	0.33	0.21

NOTE:

$\underline{\alpha}_1^Q$ =Significance level used when inverting ADF-GLS to obtain lower bound of confidence interval for ρ for left tailed Q^{GLS} test.

$\bar{\alpha}_1^Q$ =Significance level used when inverting ADF-GLS to obtain upper bound of confidence interval for ρ for right tailed Q^{GLS} test.

$\underline{\alpha}_1^t$ =Significance level used when inverting ADF-OLS for left tailed t^{OLS} test.

$\bar{\alpha}_1^t$ =Significance level used when inverting ADF-OLS for right tailed t^{OLS} test.

ξ =Scaling applied to $\bar{\alpha}_1^Q$ and $\bar{\alpha}_1^t$ in constituent Q^{GLS} and t^{OLS} tests, respectively, when applying right tailed U and W_i tests, $i = 1, 2$.

CY propose the following (infeasible) statistic for testing the null $\beta = \beta_0$:

$$Q(\beta_0, \rho) := \frac{\sum_{t=1}^T x_{t-1}^\mu \left[r_t - \beta_0 x_{t-1} - \frac{\sigma_{ue}}{\sigma_e \omega_v} (x_t - \rho x_{t-1}) \right] + \frac{T}{2} \frac{\sigma_{ue}}{\sigma_e \omega_v} (\omega_v^2 - \sigma_v^2)}{\sqrt{\sigma_u^2 (1 - \delta^2) \sum_{t=1}^T (x_{t-1}^\mu)^2}}$$

$$= Q(0, \rho) - \beta_0 \sqrt{1 / \{s^2 (1 - \delta^2)\}}$$

where $s^2 := \sigma_u^2 / \sum_{t=1}^T (x_{t-1}^\mu)^2$, σ_v^2 denotes the short run variance of v_t and $x_{t-1}^\mu := x_{t-1} - T^{-1} \sum_{s=1}^T x_{s-1}$. A confidence interval for β can then be derived based on the quantity $Q(\beta, \rho)$. Under the assumptions of CY, $Q(\beta, \rho)$ follows a standard normal distribution, resulting in the $100(1 - \alpha)\%$ confidence interval $[\underline{\beta}(\rho, \alpha), \bar{\beta}(\rho, \alpha)]$ where

$$\underline{\beta}(\rho, \alpha) = \{Q(0, \rho) + z_{\alpha/2}\} s \sqrt{1 - \delta^2},$$

$$\bar{\beta}(\rho, \alpha) = \{Q(0, \rho) - z_{\alpha/2}\} s \sqrt{1 - \delta^2}$$

with $z_{\alpha/2}$ denoting the $\alpha/2$ quantile of the standard normal distribution. To overcome the fact that $\rho = 1 - c/T$ is unknown and c cannot be consistently estimated, CY propose using a confidence interval for ρ obtained by inverting the quasi-GLS demeaned ADF-GLS t -ratio based unit root test of Elliott, Rothenberg, and Stock (1996) applied to x_t (allowing for $p - 1$ lagged difference terms, as per Assumption 1), using pre-computed (asymptotic) confidence belts. In order to prevent the resulting confidence interval for β from suffering excess coverage, CY further propose a refinement whereby the significance level used to obtain the confidence interval for ρ is adapted to upper and lower bounds separately, and also according to the

value of δ . Values of this significance level are chosen numerically to minimize over-coverage associated with the confidence interval for β , while ensuring that the overall Bonferroni test size does not exceed a chosen level across a specified range of c . Denoting the significance levels for the lower and upper confidence bounds for ρ by $\underline{\alpha}_1^Q$ and $\bar{\alpha}_1^Q$, respectively, the confidence interval for ρ can be written as $[\underline{\rho}(\underline{\alpha}_1^Q), \bar{\rho}(\bar{\alpha}_1^Q)]$, and the resulting $100(1 - \alpha_2)\%$ confidence interval for β is obtained as $[\underline{\beta}(\bar{\rho}(\bar{\alpha}_1^Q), \alpha_2), \bar{\beta}(\underline{\rho}(\underline{\alpha}_1^Q), \alpha_2)]$ where

$$\underline{\beta}(\bar{\rho}(\bar{\alpha}_1^Q), \alpha_2) = \{Q(0, \bar{\rho}(\bar{\alpha}_1^Q)) + z_{\alpha_2/2}\} s \sqrt{1 - \delta^2}, \quad (4)$$

$$\bar{\beta}(\underline{\rho}(\underline{\alpha}_1^Q), \alpha_2) = \{Q(0, \underline{\rho}(\underline{\alpha}_1^Q)) - z_{\alpha_2/2}\} s \sqrt{1 - \delta^2}. \quad (5)$$

For a given value of δ the one-sided tests for predictability constructed in this manner will have an asymptotic size of exactly $\alpha_2/2$ for some value of c while remaining slightly undersized for other values of c . Consequently, two-sided tests will have size of at most α_2 across the specified range of c . CY calibrate this procedure by fixing $\alpha_2 = 0.1$ and considering $c \in [-5, 50]$ such that their resulting one-sided tests have a maximum (asymptotic) size of 5%. The appropriate values of $\underline{\alpha}_1^Q$ and $\bar{\alpha}_1^Q$ are reproduced in Table 1 for convenience. We will denote the predictability test based on this confidence interval as Q^{GLS} in what follows.

Remark 3.1. The appropriate values of $\underline{\alpha}_1^Q$ and $\bar{\alpha}_1^Q$ reported in Table 1 are only provided for $\delta < 0$. For $\delta > 0$, CY note that replacing x_t in (1) with $-x_t$ flips the sign of both β and δ . Therefore, an equivalent right (left) tailed test for predictability when $\delta > 0$ can be performed as a left (right) tailed test for predictability based on (1) with x_t replaced by $-x_t$ using the values of $\underline{\alpha}_1^Q$ and $\bar{\alpha}_1^Q$ appropriate for a negative value of δ . This also holds for the Bonferroni t test discussed below.

Remark 3.2. If $w_0 = 0$, the infeasible test based on $Q(0, \rho)$ is, under Gaussianity, a (conditionally) uniformly most powerful test where ρ is *known* (see CY, p.32). However, the corresponding feasible Bonferroni-based Q^{GLS} test does not possess any formal optimality property. Imposing the assumption that $w_0 = 0$, EMW propose a near-optimal test for a strongly persistent predictor which is based on a weighted average (local asymptotic) power criterion. EMW impose the additional condition, not required for Q^{GLS} , that $c \geq 0$ (equivalently, $\rho \leq 1$) disallowing locally explosive predictors. CY p.54 provide a discussion on why it might not be sensible to restrict c to be nonnegative; indeed many of the predictors in CY's empirical application have confidence intervals for ρ that contain values above 1. EMW find that their test generally displays higher asymptotic local power than the CY test, although the converse can be true particularly for the case of most practical interest where δ is large and negative. Harvey, Leybourne, and Taylor (2021) find that the EMW test has uncontrolled size, even asymptotically, when $c < 0$. A version of the EMW test allowing for a locally explosive predictor would therefore likely have much lower power than the CY test.

The CES approach is based on the estimator of β obtained from OLS estimation of (1), denoted $\hat{\beta}$, and the corresponding (infeasible) t -statistic for testing the null $\beta = \beta_0 = 0$:

$$t := \frac{\hat{\beta}}{\sqrt{\sigma_u^2 / \sum_{t=1}^T x_{t-1}^{\mu 2}}}.$$

The limiting null distribution of t in the local-to-unity setting is a function of the unknown parameter c , but CES propose construction of a confidence interval for β by making use of confidence intervals for c obtained by inverting the standard OLS-demeaned ADF-OLS t -ratio based unit root test (again allowing for $p - 1$ lagged differences), again using pre-computed confidence belts. Specifically, for a given value of δ , the CES $100(1 - \alpha_2)\%$ confidence interval for β is obtained as $[\underline{\beta}(\underline{\alpha}_1^t, \alpha_2), \bar{\beta}(\underline{\alpha}_1^t, \alpha_2)]$ where

$$\underline{\beta}(\underline{\alpha}_1^t, \alpha_2) = \hat{\beta} - \left\{ \max_{c(\underline{\alpha}_1^t) \leq c \leq \bar{c}(\underline{\alpha}_1^t)} d_{c, 1-\alpha_2/2} \right\} s, \quad (6)$$

$$\bar{\beta}(\underline{\alpha}_1^t, \alpha_2) = \hat{\beta} - \left\{ \min_{c(\underline{\alpha}_1^t) \leq c \leq \bar{c}(\underline{\alpha}_1^t)} d_{c, \alpha_2/2} \right\} s \quad (7)$$

and where $d_{c, \eta}$ denotes the η -level critical value of the limiting null distribution of t for a given value of c . The significance levels used to obtain the c confidence intervals, $\underline{\alpha}_1^t$ and $\bar{\alpha}_1^t$, are selected numerically to ensure that the implied one-sided tests for predictability constructed in this manner will have an asymptotic size of exactly $\alpha_2/2$ for some value of $c \in [-5, 50]$ while remaining slightly undersized for other values of c . For $\alpha_2 = 0.1$, the appropriate values of $\underline{\alpha}_1^t$ and $\bar{\alpha}_1^t$ are those of CY, and are reported in Table 1. We will denote the predictability test based on this confidence interval as t^{OLS} in what follows.

For full details on the practical implementation of the Q^{GLS} and t^{OLS} procedures, including consistent estimation of the parameters σ_e , σ_u , σ_v , σ_{ue} , ω_v , and δ , implementation of the ADF-GLS and ADF-OLS unit root tests, and the pre-computed confidence belts, see CY, CES, and the corresponding

supplementary material to CY available at <https://scholar.harvard.edu/campbell/publications/implementing-econometric-methods-efficient-tests-stock-return-predictability-0>. The confidence belts and code for the procedures are available from Motohiro Yogo's website: <https://sites.google.com/site/motohiroyogo/research/asset-pricing>.

3.2. Asymptotic Behavior

We now consider the large sample behavior of the Q^{GLS} and t^{OLS} tests when Assumption S.3 holds, that is, the case where the predictor is a strongly persistent near-integrated process with $c > 0$ and an initial condition that is of $O_p(T^{1/2})$. In this case the limiting distributions of the statistics will be shown to depend on the magnitude of the initial condition. We will quantify this dependence, investigating the impact of the initial condition on the asymptotic size and local power of the tests.

The first step in doing so is to establish the limiting distributions of the statistics $Q(0, \tilde{\rho})$ and t , where $\tilde{\rho} = 1 - \tilde{c}/T$ for an arbitrary \tilde{c} . These results are presented in Proposition 1.

Proposition 1. Let the data on (r_t, x_t) be generated according to (1)–(3). Let $W_{e,c}(s)$ be a standard Ornstein-Uhlenbeck process on $[0, 1]$ defined by the stochastic differential equation $dW_{e,c}(s) = -cW_{e,c}(s)ds + dW_e(s)$, with initial condition $W_{e,c}(0) = 0$, and where $W_e(s)$ is a standard Weiner process. If Assumptions 1 and S.3 hold then under the local alternative $H_b : \beta = T^{-1}b$,

$$(a) \quad Q(0, \tilde{\rho}) \xrightarrow{w} \frac{b\omega_v\kappa_c^\theta}{\sigma_u\sqrt{1-\delta^2}} + \frac{\delta(c-\tilde{c})\kappa_c^\theta}{\sqrt{1-\delta^2}} + Z \quad (8)$$

$$(b) \quad t \xrightarrow{w} \frac{b\omega_v\kappa_c^\theta}{\sigma_u} + \delta \frac{\tau_c^\theta}{\kappa_c^\theta} + \sqrt{1-\delta^2}Z := t^\infty \quad (9)$$

where “ \xrightarrow{w} ” denotes weak convergence, $\kappa_c^\theta := (\int_0^1 K_{c,\theta}^\mu(s)^2 ds)^{1/2}$ and $\tau_c^\theta := \int_0^1 K_{c,\theta}^\mu(s) dW_e(s)$ with $K_{c,\theta}^\mu(s) := K_{c,\theta}(s) - \int_0^1 K_{c,\theta}(r) dr$ and

$$K_{c,\theta}(r) := \theta(e^{-rc} - 1)(2c)^{-1/2} + W_{e,c}(r) \quad (10)$$

and where $Z := (\int K_{c,\theta}^\mu(r)^2 dr)^{-1/2} \int K_{c,\theta}^\mu(s) d\tilde{W}_u(s) \sim N(0, 1)$, where $\tilde{W}_u(\cdot)$ is a standard Wiener process distributed independently of $W_e(\cdot)$.

Remark 3.3. Observe that when $\theta = 0$, $K_{c,\theta}(r)$ in (10) reduces to $W_{e,c}(r)$ and hence the limiting distributions for $Q(0, \tilde{\rho})$ and t given in Proposition 1 under Assumption S.3 simplify to the limits given in CY and CES under Assumption S.2. It follows, therefore, that asymptotic analysis of the tests under the Assumption S.2 case of $w_0 = o_p(T^{1/2})$ can be subsumed under the Assumption S.3 case of $w_0 = O_p(T^{1/2})$, on setting $\theta = 0$. Similarly, the limiting distributions of $Q(0, \tilde{\rho})$ and t under the $c = 0$ case of Assumption S.1 can also be obtained from Proposition 1 by replacing $K_{0,\theta}(r)$ with $W_e(r)$.

Remark 3.4. Where $\theta \neq 0$ it is seen from the representations in (8) and (9) that for near-integrated predictors with $c > 0$ the asymptotic distributions of both the $Q(0, \tilde{\rho})$ and t statistics

depend on θ under both the null hypothesis and local alternatives. Where the initial condition is fixed ($\sigma_\theta^2 = 0$) it follows, using the arguments made on p. 102 of Harvey and Leybourne (2005), that these limiting distributions are invariant to the sign of μ_θ .

Remark 3.5. Representations for the limiting null distributions of the $Q(0, \tilde{\rho})$ and t statistics obtain on setting $b = 0$ in the expressions in (8) and (9), respectively. Notice, therefore, that $Q(0, \rho)$ (the Q statistic calculated at the true ρ) has a standard normal limiting null distribution, although its asymptotic local power function does still depend on θ . Notice also that the local power offset in the limit of $Q(0, \tilde{\rho})$ is independent of the value of \tilde{c} .

We now evaluate the local asymptotic power of the Q^{GLS} and t^{OLS} tests under Assumptions S.1–S.3. To do so we will additionally require the limiting distributions of ADF-GLS and ADF-OLS. These are given by (see, e.g., Harvey, Leybourne, and Taylor 2009):

$$\text{ADF-GLS} \xrightarrow{w} \frac{K_{c,\theta}(1)^2 - 1}{2\sqrt{\int_0^1 K_{c,\theta}(s)^2 ds}}, \quad (11)$$

$$\text{ADF-OLS} \xrightarrow{w} \frac{K_{c,\theta}^\mu(1)^2 - K_{c,\theta}^\mu(0)^2 - 1}{2\kappa_c^\theta}. \quad (12)$$

The method for simulating the local asymptotic power proceeds as follows, where we outline the procedure for the illustrative case of right tailed testing; left tailed testing proceeds in the same manner with the obvious modifications. All simulations of limiting distributions we report were performed in Gauss 8.0 using direct simulation with 5000 Monte Carlo replications, with Wiener processes approximated using NIID(0,1) random variates and integrals approximated by normalized sums of 1000 steps.

For Q^{GLS} , we first simulate draws from the limiting distributions of ADF-GLS using (11). These values are then used to obtain the lower bound of the confidence interval for c , which we denote $\underline{c}(\bar{\alpha}_1^Q)$, using the pre-computed confidence belts discussed in Section 3.1, implemented using the values of $\bar{\alpha}_1^Q$ appropriate for δ obtained from Table 1. Note that this value of c corresponds to the upper bound of the confidence interval for ρ , that is, $\bar{\rho}(\bar{\alpha}_1^Q) = 1 - \underline{c}(\bar{\alpha}_1^Q)/T$. Of course, testing in the right tail is equivalent to determining whether $\beta(\bar{\rho}(\bar{\alpha}_1^Q), \alpha_2) > 0$, and the asymptotic local power function associated with $Q(0, \bar{\rho}(\bar{\alpha}_1^Q))$ is given by $E[\Phi(h(\bar{\alpha}_1^Q, \alpha_2))]$ where $\Phi(\cdot)$ denotes one minus the standard normal cdf and

$$h(\bar{\alpha}_1^Q, \alpha_2) := z_{1-\alpha_2/2} - \frac{b\omega_v\kappa_c^\theta}{\sigma_u\sqrt{1-\delta^2}} - \frac{\delta(c - \underline{c}(\bar{\alpha}_1^Q))\kappa_c^\theta}{\sqrt{1-\delta^2}}. \quad (13)$$

Next we simulate a draw from κ_c^θ and construct $h(\bar{\alpha}_1^Q, \alpha_2)$ in (13). Finally, we evaluate whether a simulated draw from a standard normal exceeds this value of $h(\bar{\alpha}_1^Q, \alpha_2)$. The limiting power is then obtained as the average of these exceedances across replications.

For t^{OLS} , in each simulation replication we first simulate a draw from the limiting distributions of ADF-OLS using (12),

and then obtain $[\underline{c}(\bar{\alpha}_1^t), \bar{c}(\bar{\alpha}_1^t)]$ using the corresponding pre-computed confidence belts for the values of $\bar{\alpha}_1^t$ appropriate for δ obtained from Table 1. Then we simulate the limit of t using the result in Proposition 1(b), and compare this with the critical value $\max_{\underline{c}(\bar{\alpha}_1^t) \leq c \leq \bar{c}(\bar{\alpha}_1^t)} d_{c,1-\alpha_2/2}$. The limiting power is again calculated as the average of these exceedances across replications. Note that the pre-computed confidence belts and Bonferroni refinement significance tables that are used here for Q^{GLS} and t^{OLS} are those designed for Assumptions S.1 and S.2.

In what follows we set $\sigma_u = \omega_v = 1$ without loss of generality (as these parameters can be consistently estimated) and employ the commonly used setting of $\delta = -0.95$. We report results for a fixed initial condition generated according to Assumption S.3 with $\theta = \mu_\theta = \{0, 1, 3\}$, covering cases of an asymptotically negligible initial condition ($\mu_\theta = 0$) and asymptotically nonnegligible initial conditions of increasing magnitude ($\mu_\theta = 1$ and $\mu_\theta = 3$). Results for random initial conditions (available on request) were found to be qualitatively similar and hence are not reported. We consider the local-to-unity values $c = \{0, 2, 5, 20\}$ (results for the additional cases $c = \{10, 50\}$, and for $\delta = -0.75$, are reported in the supplementary appendix) with local power curves generated across a grid of 50 values of b from 0 to a relevant value that depends on c and whether right or left tailed tests are being conducted. Recall that when $c = 0$, the tests are exact invariant to w_0 and so only one set of power results is required. Additional asymptotic size simulations covering $c = \{0, 2, 5, 10, 20, 50\}$, $\delta \in \{-0.95, -0.75, -0.50, -0.25\}$ and $\mu_\theta \in \{0, 1, 3\}$ can be found in Tables A.1–A.3 in the supplementary appendix. All tests are performed as one-sided (asymptotic) 5% tests. Results are only reported for $\delta < 0$ as the size and power for right (left) tailed tests for predictability when $\delta > 0$ are identical to left (right) tailed tests for predictability when $\delta < 0$; see Remark 3.1.

3.3. Asymptotic Size and Local Power of Right Tailed Tests when $\delta < 0$

Figure 2 graphs the asymptotic size and local power of the right tailed Bonferroni-based tests for predictability; also graphed are the corresponding quantities for the right-tailed hybrid tests that we will subsequently develop in Section 4 (discussion of which we will defer until Section 4.6). When $c = 0$, the results in panel (a) show that neither test's power profile dominates the other across all b . Consider next the case where $c > 0$ and $\mu_\theta = 0$, such that the initial condition is asymptotically negligible. It is apparent from the results in panels (b), (e), and (h) that in this scenario both tests are asymptotically size-controlled (as expected) and that the best overall local power performance is displayed by Q^{GLS} . The local power of this test offers substantial power gains relative to t^{OLS} for $c = 2, 5$, and only ever falls very slightly below that of t^{OLS} for small values of b when $c = 20$. The additional results for $c = \{10, 50\}$ reported in the supplement (see Figure A.1) show that Q^{GLS} is again arguably the best procedure, unless $c = 50$ where it lacks power relative to t^{OLS} .

We next turn our attention to the case where the initial condition of the predictor is asymptotically nonnegligible, with

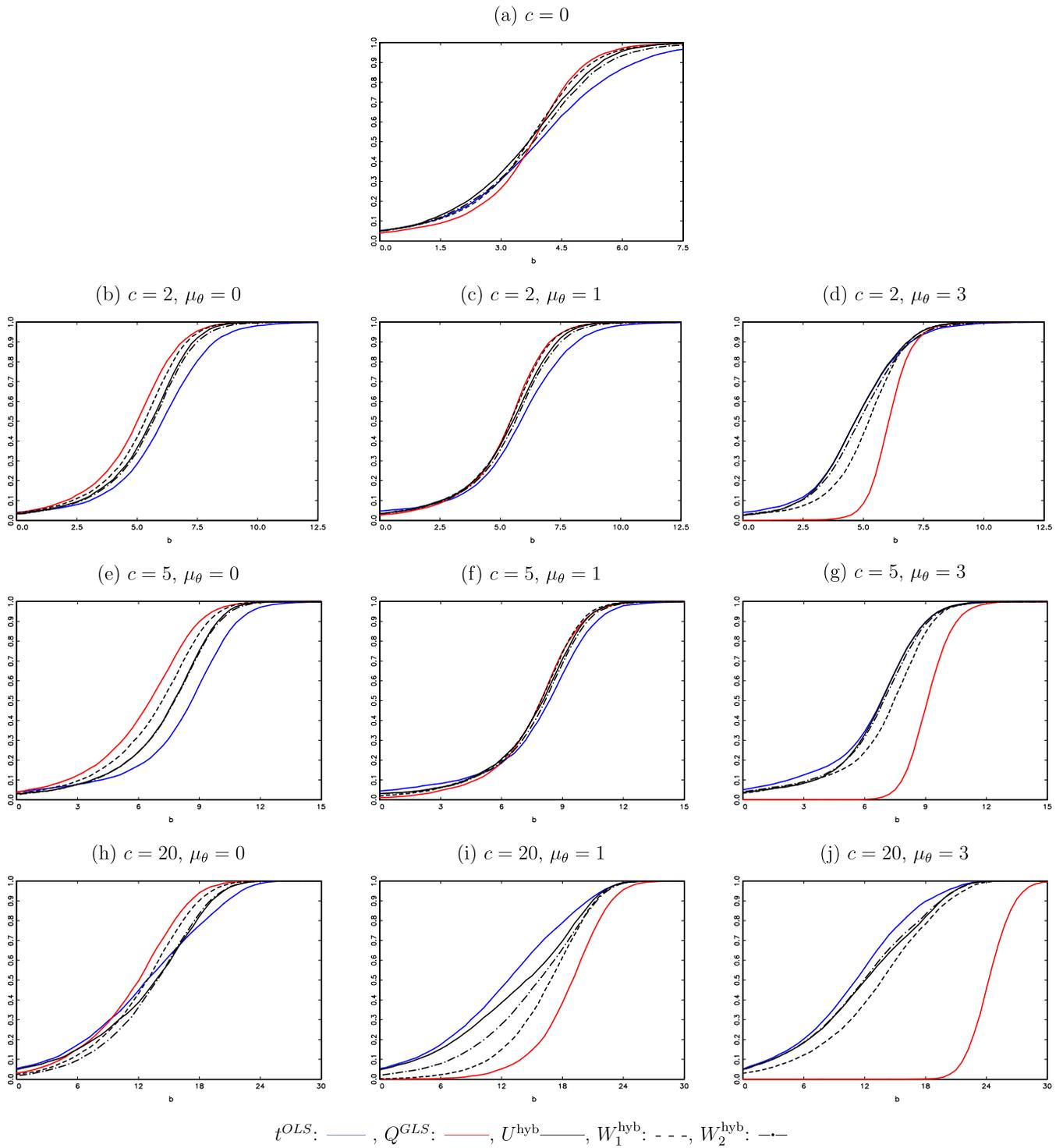


Figure 2. Asymptotic local power of right tailed tests, $\delta = -0.95$.

$\mu_\theta = 1$. We see from the results in panels (c), (f), and (i) of Figure 2 that both tests remain asymptotically size-controlled for an initial condition of this magnitude, but we observe a reduction in local power for Q^{GLS} relative to the case of $\mu_\theta = 0$, with this effect more pronounced the greater is the value of c . Indeed, the power of t^{OLS} falls only slightly below that of Q^{GLS} for $c = 2, 5$, while greatly exceeding it for the larger value of $c = 20$. Overall, the best local power performance across all values of c considered in this case is that associated with t^{OLS} .

Turning to panels (d), (g), and (j) of Figure 2 we see that a larger initial condition, with $\mu_\theta = 3$, induces a dramatic reduction in asymptotic local power for Q^{GLS} , with this test exhibiting severe undersize and a power profile that is far below that of t^{OLS} . The better overall performance for $\mu_\theta = 3$ is clearly seen to be displayed by t^{OLS} which is asymptotically size controlled and avoids the extreme under sizing seen with Q^{GLS} when the initial condition is large, and subsequently displays by far the better overall local power profile. Finally, we also note

that from the additional results in the supplementary appendix for $\mu_\theta = 1, 3$ and $c = 10, 50$, similar comments apply, lending further support to t^{OLS} being the preferred test for larger initial conditions (see Figures A.2 and A.3).

In summary, if performing right-tailed tests for predictability with $\delta < 0$ when using strongly persistent data, one should ideally perform the Q^{GLS} test when $\mu_\theta = 0$, while for larger μ_θ , the t^{OLS} test is preferable. This is an important observation that will subsequently guide the construction of our proposed hybrid predictability tests. Note that the same comments apply to left

tailed tests for predictability with strongly persistent data and $\delta > 0$, given the equivalences between the procedures discussed in Remark 3.1.

3.4. Asymptotic Size and Local Power of Left Tailed Tests when $\delta < 0$

Figure 3 reports the asymptotic size and local power of left tailed tests for predictability. In panel (a), where $c = 0$, it is clear that Q^{GLS} is the better performing test. However, when $c > 0$ with

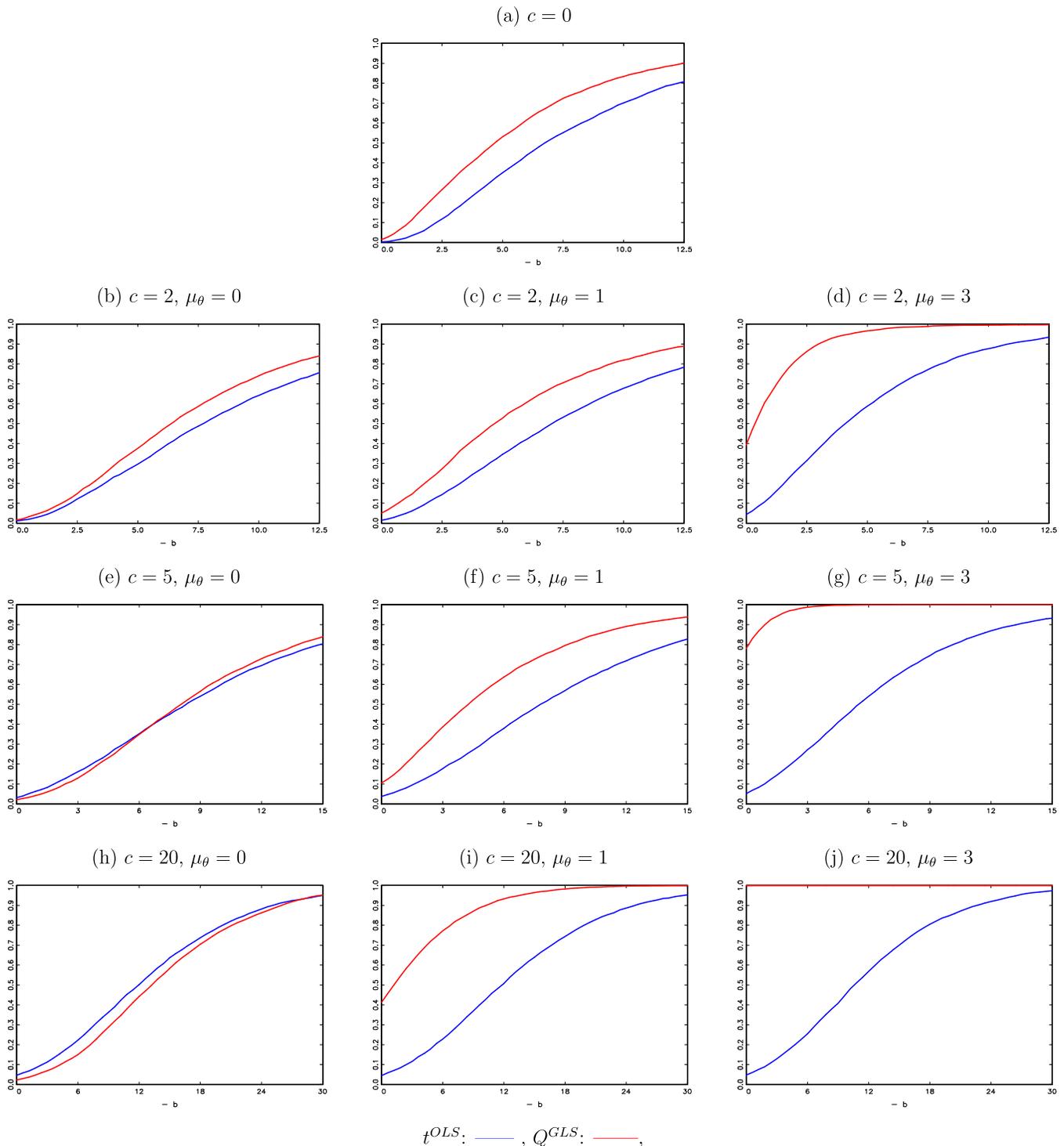


Figure 3. Asymptotic local power of left tailed Bonferroni t and Q tests, $\delta = -0.95$.

$\mu_\theta = 0$, we see from panels (b), (e), and (h) that the local power of Q^{GLS} dominates that of t^{OLS} only for $c = 2$, with the power of this test subsequently beginning to fall below that of t^{OLS} as the value of c increases. Turning to the cases where the initial condition of the predictor is asymptotically nonnegligible, it is immediately apparent that the Q^{GLS} test is not at all suitable, with significant asymptotic oversize displayed, increasingly so as both c and μ_θ increase. Here, the better performing test is t^{OLS} , with an attractive local power profile displayed across the scenarios considered.

In summary, if performing left tailed tests for predictability with $\delta < 0$ when using strongly persistent data, one should perform the t^{OLS} test, with this test displaying the better overall asymptotic size and local power profile across the scenarios considered. While the Q^{GLS} test can have local power above that of the t^{OLS} test for $c = 0$ and the smaller values of $c > 0$ when $\mu_\theta = 0$, the (often severe) oversize of this test when $c > 0$ and $\mu_\theta \neq 0$ renders it of little use empirically in this testing scenario. The same comments apply to right tailed tests for predictability with strongly persistent data and $\delta > 0$.

4. Hybrid Tests

It is clear from the previous section that, under strong persistence when $\delta < 0$, t^{OLS} should always be used for left tailed testing. The situation is, however, more complicated for right tailed testing. Here Q^{GLS} is the best overall test among those considered when $c > 0$ and $\theta = 0$, while t^{OLS} is better for larger θ when $c > 0$, with little to choose between them when $c = 0$. To that end, for right tailed testing, we first propose combined tests for predictability, designed to exploit the superior performance of the Q^{GLS} and t^{OLS} tests for different initial value magnitudes when $c > 0$. We will then discuss our full hybrid tests which switch to a standard t test if there is sufficient evidence the predictor is weakly stationary.

4.1. A Union-of-Rejections Strategy

Our first proposed combined testing procedure follows the approach taken in the context of unit root testing by Harvey, Leybourne, and Taylor (2009), and is based on a union-of-rejections strategy. Here we reject the null hypothesis of $\beta = 0$ in favor of the alternative hypothesis that $\beta > 0$ if *either* of the Q^{GLS} or t^{OLS} tests reject in the right tail. This strategy is designed to capture the excellent power properties of the Q^{GLS} test when $c > 0$ and $\theta = 0$, and the superior size and power properties of the t^{OLS} test when $c > 0$ and θ is large.

A simple union-of-rejections test based on setting $\alpha_2 = 0.1$ in connection with both of the Q^{GLS} and t^{OLS} tests was found to have a maximum asymptotic size in excess of 5% for some values of c and θ , as would be expected given that the procedure is combining rejections from two tests that are not perfectly correlated and that the calibration for the tests of CY and CES are based on the assumption that $\theta = 0$.

For a union-of-rejections test to have maximum asymptotic size of $\alpha_2/2$ we therefore need to modify the significance levels at which the initial confidence belts for ρ are constructed for both the ADF-GLS and ADF-OLS tests. Recalling that the lower

bound of the confidence interval for β obtained from the Q^{GLS} and t^{OLS} tests are given by $\underline{\beta}(\bar{\rho}(\bar{\alpha}_1^Q), \alpha_2)$ and $\underline{\beta}(\bar{\alpha}_1^t, \alpha_2)$, respectively, then our proposed union-of-rejections test, U , is formally defined by the decision rule:

$$U : \text{Reject } H_0 \text{ if } \underline{U} := \max\left(\underline{\beta}(\bar{\rho}(\xi \bar{\alpha}_1^Q), \alpha_2), \underline{\beta}(\xi \bar{\alpha}_1^t, \alpha_2)\right) > 0. \quad (14)$$

Here ξ is a scaling parameter ($\xi < 1$) chosen such that, for a given value of δ , the asymptotic size of U is no greater than $\alpha_2/2$ across a specified range of values of c and initial conditions. The local limiting behavior of \underline{U} will be detailed in Section 4.3.

4.2. A Data-Based Weighting Strategy

The union-of-rejections approach outlined above is designed to capture the desirable properties of Q^{GLS} when $\theta = 0$ and those of t^{OLS} when θ is large. This approach, however, does not incorporate any information from the sample data relating to the magnitude of the initial condition and essentially places equal weight on Q^{GLS} and t^{OLS} . One way in which sample information can be incorporated is to form a test based on a statistic which is constructed as a data-based weighted average of Q^{GLS} and t^{OLS} , where the weights used are functions based around an estimate of the initial condition (relative) magnitude, θ . The weighted tests we will propose are not based on an optimal choice of weights; doing so is infeasible in practice as it would effectively require knowledge of c and θ .³

Basing the weights on a data-based estimate of the initial condition has been shown to work well in the unit root testing context by Harvey and Leybourne (2005). They construct a test where greater weight is placed on the ADF-GLS (ADF-OLS) test when θ is estimated to be small (large).⁴ They propose using the following estimate of $|\theta|$,

$$|\hat{\theta}| := |x_0 - \hat{\mu}|/\hat{\sigma}_w \quad (15)$$

where $\hat{\mu} := T^{-1} \sum_{t=1}^T x_t$ and $\hat{\sigma}_w^2 := T^{-1} \sum_{t=1}^T (x_t - \hat{\mu})^2$. Under Assumption S.3, Harvey and Leybourne (2005, p. 102) show that $|\hat{\theta}|$ is not consistent for $|\theta|$ but has a well-defined limiting distribution that depends only on c and θ . However, based on simulating the limiting distribution of $|\hat{\theta}|$, Harvey and Leybourne (2005) argue that a monotonic relationship holds between $|\hat{\theta}|$ and $|\theta|$ so that, other things being equal, high (low)

³It might be possible, though beyond the scope of this article, to extend the EMW statistic discussed in Remark 3.2 to a family of such statistics based on maximizing weighted average power over different initial conditions, similarly to what is done in the unit root testing context by Müller and Elliott (2003). Based on this family of unit root statistics, Elliott and Müller (2006) develop an (asymptotically) admissible unit root test that, for a given value of c , has roughly constant power over a wide range of initial conditions. The predictability testing problem considered here is, however, more complicated as θ features in the limit null distributions of predictability statistics, while it does not for unit root statistics. The asymptotic size of such EMW-type tests would therefore need to be controlled over both c and θ .

⁴Harvey, Leybourne, and Taylor (2009) show that both the weighted unit test of Harvey and Leybourne (2005) and a union of rejections test based on the ADF-OLS and ADF-GLS statistics have superior asymptotic local power to the test of Elliott and Müller (2006) for either large or small initial conditions, whereas the latter is more powerful for intermediate sized initial conditions. Finite sample simulations in Harvey, Leybourne, and Taylor (2009) suggest the Elliott and Müller (2006) test is badly undersized with correspondingly poor power.

values of $|\hat{\theta}|$ are associated with high (low) values of $|\theta|$. As such, $|\hat{\theta}|$ embodies fundamental information about $|\mu_\theta|$ in the fixed initial value case and σ_θ in the random case. Using $|\hat{\theta}|$, Harvey and Leybourne (2005) propose use of the following weight function

$$\lambda_\gamma(|\hat{\theta}|) := \exp(-\gamma|\hat{\theta}|) \quad (16)$$

where $\gamma > 0$ is a user-chosen parameter. This function has the property that, for a given value of γ , as $|\hat{\theta}|$ increases in magnitude, so $\lambda_\gamma(|\hat{\theta}|)$ moves closer to zero, while as $|\hat{\theta}|$ approaches zero, so $\lambda_\gamma(|\hat{\theta}|)$ approaches one. We can, in a similar manner, make use of this $\lambda_\gamma(|\hat{\theta}|)$ function to construct a weighted average of the information from the Q^{GLS} and t^{OLS} tests. Specifically, our proposed weighted test is defined by the decision rule

$$W_\gamma : \text{Reject } H_0 \text{ if } \underline{W}_\gamma > 0 \quad (17)$$

where \underline{W}_γ is a weighted average of the confidence interval bounds associated with Q^{GLS} and t^{OLS} , viz:

$$\underline{W}_\gamma := \lambda_\gamma(|\hat{\theta}|)\underline{\beta}(\bar{\rho}(\xi\bar{\alpha}_1^Q), \alpha_2) + (1 - \lambda_\gamma(|\hat{\theta}|))\underline{\beta}(\xi\bar{\alpha}_1^t, \alpha_2). \quad (18)$$

In (18), the parameter $\xi < 1$ again allows us to control the maximum asymptotic size of W_γ in (17) at some desired level, $\alpha_2/2$.

4.3. Asymptotic Behavior of \underline{U} and \underline{W}_γ under Strong Persistence

In Corollary 1, whose proof follows directly from Proposition 1, we now detail the local limiting behavior of \underline{U} and \underline{W}_γ under Assumption S.3. The corresponding limiting behavior of \underline{U} and \underline{W}_γ under Assumptions S.1 and S.2 can also be obtained from these representations, as detailed in Remark 3.3.

Corollary 1. Let data be generated according to (1)–(3) and let Assumptions 1 and S.3 hold. Then under the local alternative $\beta = b/T$,

$$\underline{U} \xrightarrow{w} \max(\underline{\beta}^{Q,\infty}, \underline{\beta}^{t,\infty}) \quad (19)$$

$$\underline{W}_\gamma \xrightarrow{w} \exp(-\gamma A_c^\theta)\underline{\beta}^{Q,\infty} + \{1 - \exp(-\gamma A_c^\theta)\}\underline{\beta}^{t,\infty} \quad (20)$$

where $A_c^\theta := (\int_0^1 K_{c,\theta}^\mu(r)^2 dr)^{-1/2} (1 - \int_0^1 K_{c,\theta}(r) dr)$ with $K_{c,\theta}^\mu(s)$ and $K_{c,\theta}(s)$ as defined in Proposition 1 and

$$\begin{aligned} \underline{\beta}^{Q,\infty} &:= Z - h(\xi\bar{\alpha}_1^Q, \alpha_2) \\ \underline{\beta}^{t,\infty} &:= t^\infty - \max_{\xi(\bar{\alpha}_1^t) \leq c \leq \bar{c}(\xi\bar{\alpha}_1^t)} d_{c,1-\alpha_2/2} \end{aligned}$$

where $h(\cdot)$ and t^∞ are as defined in (13) and (9), respectively, and Z is a standard normal random variable that is independent of $W_e(s)$.

Remark 4.1. As noted above, \underline{W}_γ is a function of γ . In what follows we will report numerical and empirical results for tests based on $\gamma = 1$ and $\gamma = 2$. Increasing γ implicitly places more weight on the t^{OLS} test relative to the Q^{GLS} test.

To control the asymptotic size of U and W_γ , $\gamma = 1, 2$, values of ξ were chosen such that the asymptotic size of each test was no

greater than 5% over a grid of values of $c \in [-5, 50]$, operating under Assumptions S.1, S.2, and S.3 for $c = 0$, $c < 0$ and $c > 0$, respectively. When $c > 0$, we further ensure asymptotic size is controlled across both fixed and random initial conditions using grids of values of $\mu_\theta \in [0, 3]$ and $\sigma_\theta \in [0, 3]$. The required ξ values, obtained by simulation of the relevant limiting distributions, are reported in Table 1.

4.4. Allowing for Weakly Persistent Predictors

Under Assumption W, such that the predictor is weakly persistent, the Bonferroni Q and t tests discussed in Section 3 are asymptotically invalid. Moreover, in our Monte Carlo exercise reported in Tables A.4–A.6 in the supplementary appendix, we find that although the finite sample size of the t^{OLS} test remains reasonably well controlled for small values of ρ , the Q^{GLS} test suffers from severe oversize in this case when testing for predictability in either tail with $\delta < 0$. The behavior of Q^{GLS} therefore renders the combined tests U and W_γ unreliable for use with weakly persistent predictors. Moreover, where the predictor is weakly persistent, the conventional regression t test using standard normal critical values is asymptotically optimal (among feasible tests) under Gaussianity; see Jansson and Moreira (2006, p. 704).

Based on the foregoing observations, we propose a hybrid testing approach, similar in spirit to that used in EMW and Harvey, Leybourne, and Taylor (2021), whereby we switch from the use of the Bonferroni-based combined tests U and W_γ to a standard t test, compared with normal critical values, if the data provide sufficient evidence that the predictor is weakly persistent. To that end, and following Harvey, Leybourne, and Taylor (2021), we propose using the Dickey-Fuller normalized bias coefficient unit root statistic, defined by $ADF_\phi := (T\hat{\phi}) / (1 - \sum_{i=1}^{p-1} \hat{\psi}_i)$, where $\hat{\phi}$ and $\hat{\psi}_i$, $i = 1, \dots, p - 1$ are obtained by OLS estimation of

$$\Delta x_t = \pi + \phi x_{t-1} + \sum_{i=1}^{p-1} \psi_i \Delta x_{t-i} + e_t. \quad (21)$$

In practice, p can be chosen by any consistent method; we use the BIC in our finite sample simulations and empirical application. Under Assumptions S.1–S.3, $ADF_\phi = O_p(1)$, while under Assumption W ADF_ϕ diverges to minus infinity at a rate faster than $T^{1/2}$. Employing any fixed critical value for ADF_ϕ would therefore ensure that, at least in large samples, the conventional t test would always be selected under weak persistence. However, use of a fixed critical value can result in the conventional t test also being selected under strong persistence. To control for this we therefore implement our switching rule with a diverging critical value, $-\kappa_\phi T^{1/2}$, $\kappa_\phi > 0$, so that the conventional t test is used whenever $ADF_\phi < -\kappa_\phi T^{1/2}$. The divergence rate of ADF_ϕ ensures that, in large samples, the conventional t test will be performed for weakly persistent predictors, while the Bonferroni type tests are performed for strongly persistent predictors. Although this decision rule is valid for any positive value of κ_ϕ , we found that a choice of $\kappa_\phi = 4.5$ led to the best overall finite sample size control for the hybrid procedures and so we set $\kappa_\phi = 4.5$ in our finite sample simulations and empirical application.

4.5. Proposed Hybrid Testing Procedures

On the basis of the preceding results, we now formally detail our two new hybrid testing procedures which we denote U^{hyb} and W_γ^{hyb} , $\gamma = 1, 2$, in what follows. We outline these based on the assumption that $\delta < 0$, with $\delta > 0$ subsequently discussed in Remark 4.3. Our proposed decision rules for one-sided tests performed at the $\alpha/2$ nominal asymptotic level can be written as follows, where we, again, denote the α quantile of the normal distribution as z_α .

Decision Rule for Hybrid Test Procedures ($\delta < 0$)

- **Right Tailed Tests:**
 - **Decision Rule for U^{hyb} :**
 - * If $ADF_\phi < -\kappa_\phi T^{1/2}$: **Reject H_0 if $t > z_{1-\alpha/2}$**
 - * If $ADF_\phi \geq -\kappa_\phi T^{1/2}$: **Reject H_0 if $\underline{U} > 0$** (see Equation 14)
 - **Decision Rule for W_γ^{hyb} :**
 - * If $ADF_\phi < -\kappa_\phi T^{1/2}$: **Reject H_0 if $t > z_{1-\alpha/2}$**
 - * If $ADF_\phi \geq -\kappa_\phi T^{1/2}$: **Reject H_0 if $\underline{W}_\gamma > 0$** (see Equation 18)
- **Left Tailed Tests:**
 - **Decision Rule for U^{hyb} and W_γ^{hyb} :**
 - * If $ADF_\phi < -\kappa_\phi T^{1/2}$: **Reject H_0 if $t < z_{\alpha/2}$**
 - * If $ADF_\phi \geq -\kappa_\phi T^{1/2}$: **Reject H_0 if $\overline{\beta}(\underline{\alpha}_1^t, \alpha_2) < 0$** (see Equation 7)

Remark 4.2. Notice that the decision rules for U^{hyb} and W_γ^{hyb} are identical for left-tailed tests when $\delta < 0$ as here inference is always based on either the Bonferroni t^{OLS} test or on a conventional regression t test.

Remark 4.3. When $\delta > 0$, we make use of the result in Remark 3.1 and suggest replacing the predictor x_t in (1) with $-x_t$, thereby flipping the sign of δ such that our recommended procedures for negative values of δ can then be applied. In this instance, however, it should be noted that the sign of β will also flip, so that if one were interested in a right (left) tailed test for predictability one should instead perform a left (right) tailed test for predictability in the transformed predictive regression that contains $-x_t$ as a regressor. In practice, the true value of δ will be unknown, but the appropriate approach can be determined according to the sign of the consistent estimator, $\hat{\delta}$.

Remark 4.4. The switching decision rule outlined above can also be applied to the original Q^{GLS} test of CY when implemented as a right-tailed test with negative δ , or as a left-tailed test with positive δ . Here one uses the Bonferroni Q^{GLS} test as outlined in CY, unless $ADF_\phi < -\kappa_\phi T^{1/2}$ in which case the conventional t test is used. This hybrid switching-based testing procedure is asymptotically valid for both weakly and strongly persistent

predictors, generated according to Assumption W or any of Assumptions S.1–S.3, respectively. It should be stressed however that this procedure would suffer from the same undersizing and low power as the original Q^{GLS} test for strongly persistent predictors with asymptotically nonnegligible initial conditions generated according to Assumption S.3. Moreover, it could not be validly implemented as a left-tailed test with negative δ , or as a right-tailed test with positive δ , because of the uncontrolled size of the Q^{GLS} test in those settings.

Remark 4.5. While the definitions of the hybrid U^{hyb} and W_γ^{hyb} procedures given above are framed in terms of one-sided tests for predictability, in principle each of these procedures can also be used to perform two-sided tests for predictability. For a given test, if the right tailed and left tailed versions of the test are constructed such that they have nominal size no greater than $\alpha/2$, then combining inference from the two individual one-sided tests for predictability will lead to an overall two-sided test for predictability that will have nominal size no greater than α .

4.6. Asymptotic Size and Local Power of Hybrid Procedures

In this section we report results of a Monte Carlo simulation study in which we examine the asymptotic size and local power of our proposed U^{hyb} and W_γ^{hyb} tests relative to the Q^{GLS} and t^{OLS} procedures. We report results for the same constellation of settings as in Section 3.2 (additional simulations exploring the case where $\delta = -0.75$, as well as a larger range of values of c , together with additional asymptotic size simulations are provided in Figures A.1–A.6 and Tables A.1–A.3 in the supplementary appendix). We place our focus on right tailed tests for predictability given that the construction of the U^{hyb} and W_γ^{hyb} procedures implies that they will have identical local asymptotic power functions to t^{OLS} when performing left tailed tests for predictability. The results are reported in Figure 2.

First, we note that when $c = 0$, the hybrid tests perform very well, being size controlled (see also Tables A.1–A.3) and arguably as powerful as any of the individual tests, with power exceeding that of the other procedures for small b and only slightly below the power of the best individual procedure for larger b .

For $c > 0$, consider first the case where $\mu_\theta = 0$ such that the initial condition is asymptotically negligible. As would be expected, the new hybrid procedures are asymptotically size-controlled across the different values of c (again see also Tables A.1–A.3). In terms of asymptotic local power, again as expected, Q^{GLS} remains the best procedure in terms of overall power across the values of c considered, with W_1^{hyb} the next best performing procedure with power only marginally lower than Q^{GLS} , while having uniformly higher power than all other procedures for $c = 2, 5$ and one of the better overall power profiles for $c = 20$. The U^{hyb} procedure has power that is overall not far behind that of W_1^{hyb} , with W_2^{hyb} displaying marginally lower power overall than U^{hyb} . Moreover, as the results of the supplementary Figure A.1 shows, by $c = 50$ the U^{hyb} procedure becomes more powerful than W_1^{hyb} (and W_2^{hyb}), being relatively unaffected by the drop off in power associated with Q^{GLS} , instead displaying the same power levels as the now better-performing t^{OLS} test.

Next we consider the asymptotically nonnegligible initial condition cases of $\mu_\theta = 1, 3$ when $c > 0$. We first observe that the three hybrid procedures retain asymptotic size control in these cases (again see also Tables A.1–A.3). For $\mu_\theta = 1$ we see that, across all values of c considered, the best hybrid procedure in terms of overall local power performance is clearly U^{hyb} , with power levels either a little greater or a little below those of t^{OLS} . While W_1^{hyb} and W_2^{hyb} display decent power performance for lower values of c , they do exhibit a significant shortfall in power relative to U^{hyb} for $c = 20$, although they are still far more powerful than Q^{GLS} in this instance. For $c = 20$ we also see that W_2^{hyb} outperforms W_1^{hyb} as anticipated, given that the former places lower weight on the less powerful Q^{GLS} test than the latter. When $\mu_\theta = 3$, U^{hyb} is again the best performing hybrid procedure, although we now see that the powers of W_1^{hyb} and W_2^{hyb} are much closer to those of U^{hyb} , particularly so for W_2^{hyb} . The U^{hyb} procedure is again competitive with the best of the individual tests, t^{OLS} , in this case. Similar comments apply to the supplemental results for the additional values of c (Figures A.2 and A.3), with the $c = 50$ case representing a more exaggerated version of the $c = 20$ results. Finally, we note that the supplementary results for $\delta = -0.75$ follow much the same pattern as for $\delta = -0.95$, albeit with the power differentials between the tests being somewhat less pronounced.

Overall, across asymptotically negligible and nonnegligible initial conditions, for right tailed tests for predictability we argue that the best overall asymptotic local power performance is displayed by the new hybrid procedure U^{hyb} , with the performance of W_1^{hyb} and W_2^{hyb} not far behind. Importantly, while the Q^{GLS} and t^{OLS} tests are arguably the best performing individual tests for $\mu_\theta = 0$ and $\mu_\theta = 1, 3$, respectively, these approaches do not deliver the best power profiles across the full range of initial condition magnitudes, with Q^{GLS} and t^{OLS} performing relatively poorly for $\mu_\theta = 1, 3$ and $\mu_\theta = 0$, respectively. The value of the new procedures is therefore clearly evident in the practical situation of dealing with a strongly persistent predictor where the magnitude of the initial condition is unknown.

Additional Monte Carlo results, reported in Tables A.4–A.6 and Figures A.7–A.18 in the supplementary appendix, under each of Assumption S.1, Assumption S.2 with $w_0 \sim N(0, 1)$, and Assumption S.3 with $\sigma_\theta^2 = 0$ and $\mu_\theta = 1, 3$, show that, for a strongly persistent predictor, the attractive large sample size and power properties of our hybrid test procedures hold even for a relatively modest sample size ($T = 250$). These simulations also show that the hybrid tests have well controlled size and strong power properties in the case where c is large ($c \geq 100$), such that the predictor may reasonably be characterized as weakly persistent, because they switch into the standard t test with sufficiently high probability to avoid the severe distortions from nominal size which occur for the Q^{GLS} test and/or moderate size distortions seen for the t^{OLS} tests.

5. Empirical Application

We now report results of an empirical exercise in which we revisit the dataset originally analyzed in CY to further illustrate the sensitivity of their Q^{GLS} test to the value of the initial

condition of the predictor, and to explore to what extent our proposed hybrid tests, U^{hyb} and W_γ^{hyb} , are able to overcome these shortcomings.

As a preliminary analysis we applied the Q^{GLS} , t^{OLS} , U^{hyb} and W_γ^{hyb} test procedures to the same empirical returns/predictor pairings considered in CY, but rather than applying the procedures to only the full sample of data, we applied them recursively across all possible start dates, t_s , subject to a minimum sample size of 50 observations. We examine predictability of returns for both the S&P500 and CRSP indices. The predictors considered are the earnings-price ratio ($e - p$), the dividend price ratio ($d - p$), the three-month T-Bill rate (r_3) and the long-short yield spread ($y - r_1$). Full data descriptions are provided in CY.⁵ All tests are performed as one-sided tests at the nominal 5% level. All unit root statistics used to construct the tests for predictability were, following CY, estimated using a lag length chosen by the BIC applied to the ADF-OLS regression with $p_{\text{max}} = 5$.

Table 2 provides a summary of these results including the full sample estimates of the correlation parameter, $\hat{\delta}$, and the Dickey-Fuller normalized bias coefficient unit root test statistic (critical value in parentheses), along with the proportion of start dates for which each test rejects the null of no predictability (entries in bold highlight the procedure with the largest proportion of rejections). Results for annual CRSP 1952–2002 are omitted as this dataset contains only 51 observations. All tests are performed as right tailed tests, excepting those for CRSP returns from 1952 to 2002 using r_3 as a predictor which are performed as left tailed tests (CY find the coefficient on the predictor in these examples to be significantly negative). It can be noted that for a majority of return/predictor pairings the estimate of the correlation parameter, $\hat{\delta}$, is found to be large and negative, adding further motivation to our choice of $\delta = -0.95$ in the simulations of Sections 3.2 and 4.6. There were only five subsample regressions in the entire exercise for which the Dickey-Fuller normalized bias statistic was found to be less than $-4.5T^{1/2}$, all of which were for quarterly CRSP data for the sample 1952–2002 with either r_3 or $y - r_1$ used as the predictor, such that our hybrid tests are performed assuming that the data is strongly persistent in a vast majority of cases.

When using data from 1880 to 2002 for the S&P500 or from 1926 to 2002 for the CRSP index we see that the Q^{GLS} test rejects marginally more often across start dates than either U^{hyb} or W_γ^{hyb} , although the reverse is often true when using S&P500 data from 1880 to 1994 or CRSP data from 1926–1994. In general the W_1^{hyb} procedure rejects more often than either U^{hyb} or W_2^{hyb} . The t^{OLS} test generally has a much lower overall rejection frequency than all other tests. There is very little difference in rejection rates between the procedures when using data on the CRSP index from 1952 to 2002, with no evidence of predictability found for any start date by any procedure when using $d - p$ or $e - p$ as a predictor. That there is little variation in rejection rates across procedures when using either r_3 or $y - r_1$ as a predictor is unsurprising given that the $\hat{\delta}$ values are close to

⁵Implementation code is available from <https://rtaylor-essex.droppages.com/esrc2/default.htm> and the data can be obtained from <https://sites.google.com/site/motohiroyogo/research/asset-pricing>. Graphs of the returns series and predictors forming the dataset are given in section A.3 of the Supplementary appendix.

Table 2. Empirical application summary.

Series	Predictor	$\hat{\delta}$	Rejection frequency					
			$ADF_{\hat{\phi}}(cv)$	t^{OLS}	Q^{GLS}	U^{hyb}	W_1^{hyb}	W_2^{hyb}
S&P: 1880-2002, CRSP: 1926-2002								
S&P 500	$d - p$	-0.845	-5.76(-49.91)	8.1%	23.0%	16.2%	20.3%	14.9%
	$e - p$	-0.962	-16.97(-49.91)	5.4%	68.9%	39.2%	56.8%	35.1%
Annual CRSP	$d - p$	-0.721	-5.22(-39.49)	67.9%	85.7%	71.4%	75.0%	75.0%
	$e - p$	-0.957	-11.17(-39.49)	14.3%	46.4%	39.3%	42.9%	25.0%
Quarterly CRSP	$d - p$	-0.942	-11.15(-78.59)	6.6%	22.3%	19.9%	21.5%	14.8%
	$e - p$	-0.986	-12.87(-78.59)	3.9%	15.2%	8.6%	14.1%	10.5%
Monthly CRSP	$d - p$	-0.950	-12.06(-135.97)	6.6%	18.9%	15.0%	17.2%	11.5%
	$e - p$	-0.987	-10.46(-135.97)	2.4%	11.8%	4.1%	8.9%	4.9%
S&P: 1880-1994, CRSP: 1926-1994								
S&P 500	$d - p$	-0.835	-14.91(-48.26)	28.8%	33.3%	22.7%	25.8%	22.7%
	$e - p$	-0.958	-23.72(-48.26)	92.4%	100.0%	98.5%	100.0%	97.0%
Annual CRSP	$d - p$	-0.693	-11.63(-37.38)	100.0%	100.0%	100.0%	100.0%	100.0%
	$e - p$	-0.959	-16.45(-37.38)	90.0%	90.0%	100.0%	100.0%	100.0%
Quarterly CRSP	$d - p$	-0.941	-19.65(-74.35)	36.2%	41.1%	36.2%	40.2%	36.2%
	$e - p$	-0.988	-18.80(-74.35)	31.7%	31.7%	35.3%	38.4%	34.4%
Monthly CRSP	$d - p$	-0.948	-21.18(-128.62)	25.7%	29.0%	29.9%	32.0%	29.2%
	$e - p$	-0.983	-18.53(-128.62)	27.3%	26.6%	30.2%	33.6%	29.6%
CRSP: 1952-2002								
Quarterly CRSP	$d - p$	-0.977	-5.39(-64.27)	0.0%	0.0%	0.0%	0.0%	0.0%
	$e - p$	-0.980	-5.59(-64.27)	0.0%	0.0%	0.0%	0.0%	0.0%
	r_3	-0.095	-10.83(-64.27)	4.5%	5.2%	4.5%	4.5%	4.5%
Monthly CRSP	$y - r_1$	-0.100	-25.92(-64.27)	50.3%	49.0%	50.3%	46.5%	46.5%
	$d - p$	-0.967	-5.32(-111.32)	0.0%	0.0%	0.0%	0.0%	0.0%
	$e - p$	-0.982	-4.87(-111.32)	0.0%	0.0%	0.0%	0.0%	0.0%
	r_3	-0.071	-12.43(-111.32)	14.6%	14.7%	14.6%	14.6%	14.6%
	$y - r_1$	-0.066	-45.70(-111.32)	48.1%	47.6%	48.1%	48.0%	48.1%

zero, with unreported simulations showing that all tests share a very similar power profile for $\delta = 0$. Consequently, we do not consider the 1952–2002 CRSP data further.

The information in Table 2 only gives us a broad overview of the behavior of the procedures when applied across various start dates, so we now revisit our empirical case study from Section 1 and provide a more detailed analysis and discussion of the variability in test rejections in relation to the value of the initial condition for the CRSP data from 1926 to 1994 for the earnings-price ratio predictor, $e - p$. Corresponding results and discussion for other returns/predictor pairings are explored in the supplementary appendix. Given that there is little evidence of $e - p$ being a significant predictor of returns in the post-war period based on the predictive regressions using the 1952–2002 CRSP data summarized in Table 2, we consider start dates, t_s , for the predictive regressions up to and including the end of 1945. For these data series, we now examine each procedure Q^{GLS} , t^{OLS} , U^{hyb} , W_1^{hyb} , and W_2^{hyb} in detail, investigating the pattern of rejections relative to the magnitude of the initial condition across start dates up to and including the end of 1945.

Figure 4 reports the lower bound of the confidence interval for β for each procedure for the quarterly CRSP returns data when using the earnings-price ratio as a predictor, with green highlights indicating rejection, and red highlights non-rejection (the grey shaded regions further highlight regions of nonrejection), with the blue line plotting $|\hat{\theta}|$. Also reported in the subfigure legends for each test is the percentage of start dates for which each procedure rejects across the range of start dates considered in the figure. For this series we see from Figure 4(a) that while the Q^{GLS} test rejects for 77% of start dates considered, there is a

large window of start dates from $t_s = 1931Q3$ through to $t_s = 1935Q1$, as well as the start dates $t_s = 1942Q1$ and $t_s = 1942Q2$, for which the Q^{GLS} test fails to reject the null of no predictability. It is clearly seen that these start dates are associated with many of the largest values of $|\hat{\theta}|$ for this predictor. As a consequence of the numerous large values of $|\hat{\theta}|$, the t^{OLS} test (Figure 4(b)) actually rejects with greater frequency than the Q^{GLS} test, although the t^{OLS} test does fail to reject for a number of later start dates where $|\hat{\theta}|$ is small. The W_1^{hyb} procedure (Figure 4(d)), on the other hand, rejects for each and every start date, and the U^{hyb} (Figure 4(c)) and W_2^{hyb} (Figure 4(e)) procedures reject for 96% and 97% of start dates, respectively, with greater consistency displayed by these procedures across the varying magnitudes of $|\hat{\theta}|$ than either the Q^{GLS} or t^{OLS} tests.

Figure 5 reports results for the monthly CRSP returns when using the $e - p$ predictor which is the empirical example we originally examined for Q^{GLS} in Figure 1 in the Introduction. For this example we observe a stark difference between Q^{GLS} and all other tests, with the overall rejection frequency of Q^{GLS} standing at 74%, that for t^{OLS} at 90%, and our proposed tests at 93% or above. The Q^{GLS} test fails to reject for two large windows of start dates from $t_s = 1928M10$ through to $t_s = 1929M9$ and $t_s = 1931M8$ through to $t_s = 1935M4$, whereas all other test procedures reject for every possible start date in these two windows (with the exception of W_1^{hyb} for $t_s = 1931M8$). In both instances these windows of start dates for which Q^{GLS} fails to reject are associated with large values of $|\hat{\theta}|$, with the longer run of nonrejections associated with a period in which $|\hat{\theta}|$ is very large indeed. While the rejection frequency for the t^{OLS} test is

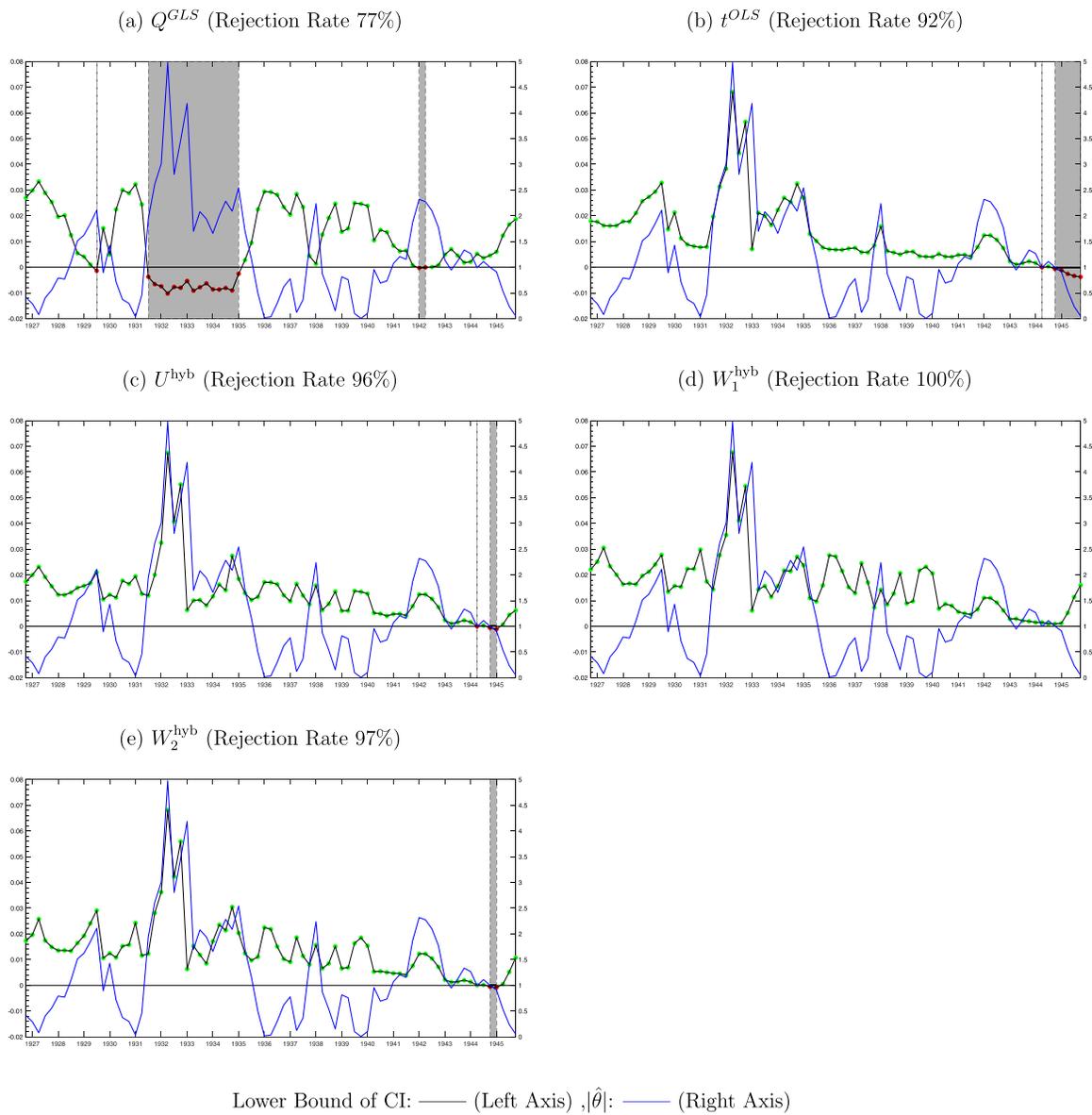


Figure 4. Lower bound of confidence interval and estimated magnitude of initial condition—Quarterly CRSP 1926-1994 (Predictor = $e - p$).

not far behind that of our proposed tests we note that t^{OLS} fails to reject for $t_s = 1930M10$ through to $t_s = 1931M4$, start dates for which all other tests continue to reject, and is also less likely to reject than all other tests for a number of later start dates. In all instances these start dates coincide with very small estimates of $|\hat{\theta}|$.

Overall our findings for these series show that the U^{hyb} and W_γ^{hyb} , $\gamma = 1, 2$, procedures reject more often than both the Q^{GLS} and t^{OLS} tests across the range of start dates considered which show large variation in an estimate of the size of the initial condition. A large initial condition can have a negative impact on the capacity for Q^{GLS} to reject, whereas a small initial condition results in t^{OLS} rejecting less frequently than the other tests. That our proposed tests are able to reject more consistently than both Q^{GLS} and t^{OLS} tallies with our asymptotic and finite sample simulation results, and reinforces our conclusion that the hybrid procedures can deliver more consistent power

across large and small magnitudes of the predictor’s initial condition.

6. Conclusions

We have demonstrated that the Bonferroni Q test of CY, while displaying excellent power when testing for predictability when a predictor is strongly persistent with an asymptotically negligible initial condition, suffers from severe size distortions and power losses when either the initial condition of the predictor is asymptotically nonnegligible or the predictor is weakly persistent. We subsequently proposed two new hybrid testing procedures, both of which are functions of the Bonferroni Q test of CY, the Bonferroni t test of CES, and the conventional t test. We have shown that the asymptotic local power of our proposed hybrid tests is close to that of the Bonferroni Q test when the initial condition is asymptotically negligible, and far superior when the

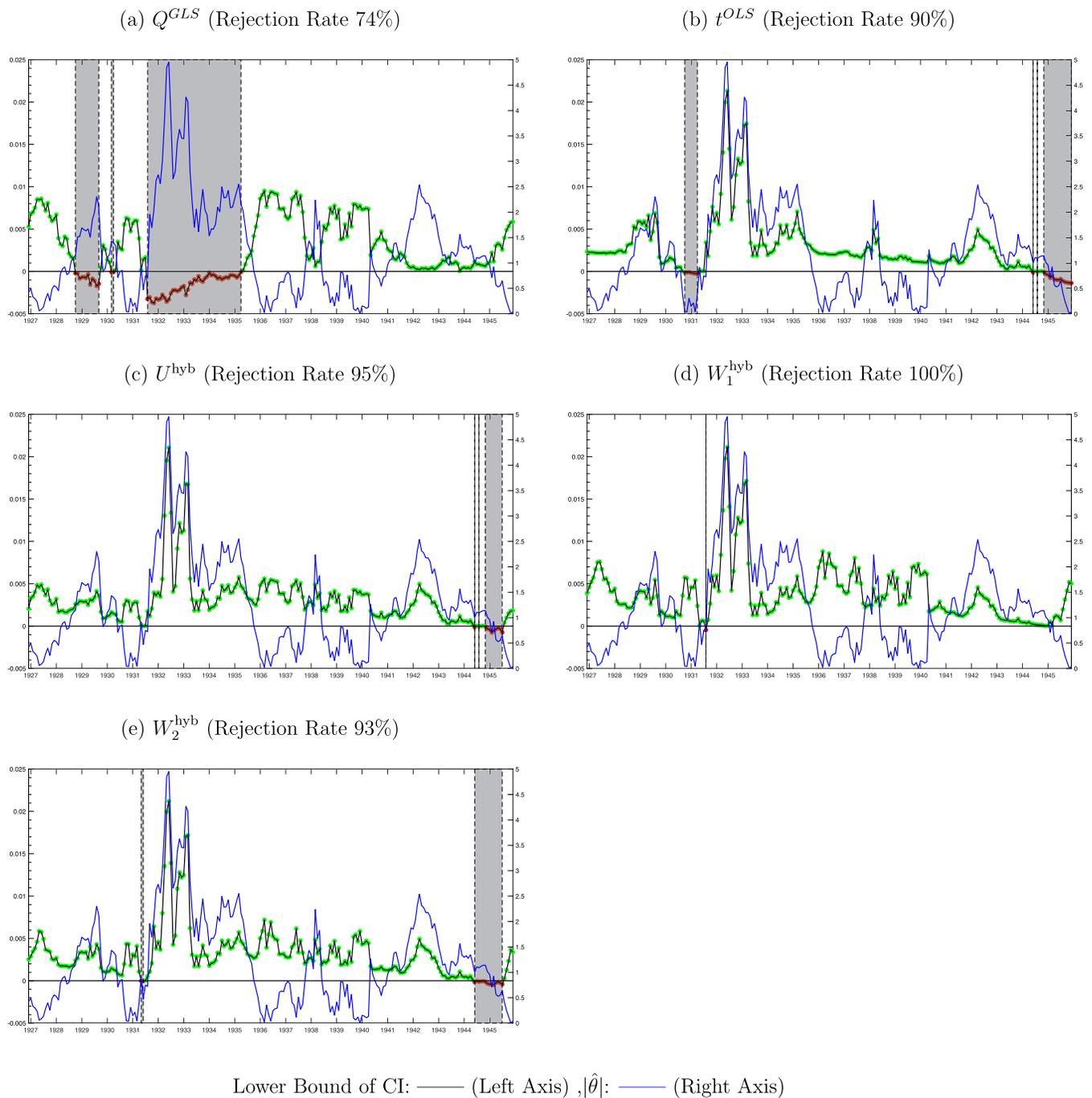


Figure 5. Lower bound of confidence interval and estimated magnitude of initial condition—Monthly CRSP 1926–1994 (Predictor = $e - p$).

initial condition is asymptotically nonnegligible. An extensive Monte Carlo simulation exercise provided in the supplementary appendix examining the finite sample size and power of the hybrid procedures shows that they are able to control size regardless of both the degree of persistence and magnitude of the initial condition of the predictor while maintaining power close to that of the Bonferroni Q test when the predictor is strongly persistent with an asymptotically negligible initial condition. An empirical application to the returns and predictor data originally analyzed in CY highlighted the ability of our proposed hybrid tests to provide statistically significant evidence of predictability where the Bonferroni Q and t tests fail to do

so in cases where the magnitude of the initial condition of the predictor is estimated to be large or small, respectively. Given that both the initial condition and the degree of persistence of a given predictor are unknown in practice we believe that our proposed hybrid testing procedures will be very useful to empirical practitioners. In particular, the loss of power of the hybrid tests relative to the Bonferroni Q test when the predictor is strongly persistent with an asymptotically negligible initial condition is very small compared to the superior size control and large power advantages displayed by the hybrid tests when the initial condition of the predictor is large or the predictor is weakly persistent.

Acknowledgments

We thank two anonymous referees, an Associate Editor, and the Coeditor, Atsushi Inoue, for their helpful and constructive comments. We are grateful to Motohiro Yogo for making his Gauss programs to implement the Bonferroni Q and t tests publicly available on his website.

Funding

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

ORCID

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

References

- Breitung, J., and Demetrescu, M. (2015), “Instrumental Variable and Variable Addition based Inference in Predictive Regressions,” *Journal of Econometrics*, 187, 358–375. [1]
- Campbell, J. Y., and Yogo, M. (2006), “Efficient Tests of Stock Return Predictability,” *Journal of Financial Economics*, 81, 27–60. [1]
- Cavanagh, C. L., Elliott, G., and Stock, J. H. (1995), “Inference in Models with Nearly Integrated Regressors,” *Econometric Theory*, 11, 1131–1147. [1]
- Elliott, G., and Müller, U. K. (2006), “Minimizing the Impact of the Initial Condition on Testing for Unit Roots,” *Journal of Econometrics* 135, 285–310. [10]
- Elliott, G., Müller, U. K., and Watson, M. W. (2015), “Nearly Optimal Tests when a Nuisance Parameter is Present Under the Null Hypothesis,” *Econometrica*, 83, 771–811. [3]
- Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996), “Efficient Tests for an Autoregressive Unit Root,” *Econometrica*, 64, 813–836. [2,5]
- Fama, E. F. (1981), “Stock Returns, Real Activity, Inflation, and Money,” *American Economic Review*, 71, 545–565. [1]
- Harvey, D. I., and Leybourne, S. J. (2005), “On Testing for Unit Roots and the Initial Observation,” *Econometrics Journal*, 8, 97–111. [2,3,7,10,11]
- Harvey, D. I., Leybourne, S. J., and Taylor, A. M. R. (2009), “Unit Root Testing in Practice: Dealing with Uncertainty over the Trend and Initial Condition,” *Econometric Theory*, 14, 587–636. [3,7,10]
- Harvey, D. I., Leybourne, S. J., and Taylor, A. M. R. (2021), “Simple Tests for Stock Return Predictability with Good Size and Power Properties,” *Journal of Econometrics*, 224, 198–214. [3,6,11]
- Jansson, M., and Moreira, M. J. (2006), “Optimal Inference in Regression Models with Nearly Integrated Regressors,” *Econometrica*, 74, 681–714. [3,11]
- Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015), “Robust Econometric Inference for Stock Return Predictability,” *Review of Financial Studies*, 28, 1506–1553. [1]
- Müller, U. K., and Elliott, G. (2003), “Tests for Unit Roots and the Initial Condition,” *Econometrica*, 71, 1269–1286. [2,3,4,10]
- Welch, I., and Goyal, A. (2008), “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction,” *Review of Financial Studies*, 21, 1455–1508. [1]