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]

[a] Essex Business School, University of Essex

[b] Granger Centre for Time Series Econometrics and School of Economics, University of Nottingham

February 14, 2022

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 (2006) 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 *et al.* (1995), 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 non-negligible. 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 analysed by Campbell and Yogo (2006) 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.

Keywords: predictive regression; initial condition; unknown regressor persistence; Bonferroni tests; hybrid tests.

JEL Classification: C22; C12; G14.

*We are grateful to Motohiro Yogo for making his Gauss programs to implement the Bonferroni Q and t tests publicly available on his website. Taylor gratefully acknowledges financial support provided by the Economic and Social Research Council of the United Kingdom under research grant ES/R00496X/1. Address correspondence to: Robert Taylor, Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK. Email: robert.taylor@essex.ac.uk.

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), for instance, 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-price ratio, the earnings-price ratio, 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 corresponding slope coefficient β.

A common finding in empirical studies into return predictability is that the putative predictor is often both highly persistent, being found to be a unit root or near unit root autoregressive process, and 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*, CY and Welch and Goyal (2008). In this situation Cavanagh *et al.* (1995) [CES] 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, modelled 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 paper.[1]

[1] Another strand of the literature characterised by contributions from Phillips and Magdalinos (2009), Kostakis *et al.* (2015) and Breitung and Demetrescu (2015) focusses on instrumental variable [IV] estimation using an instrument constructed from the predictor variable that is designed to be less persistent than a local-to-unity process. While such IV based tests are valid regardless of whether the predictor is weakly or strongly persistent, they are less powerful than the Q test when the predictor is strongly persistent, a significant drawback considering a large number of candidate predictors in empirical work appear to be strongly persistent. Other alternative hybrid approaches designed to be robust to the degree of persistence in the predictor are developed by Elliott *et al.* (2015) and Harvey *et al.* (2021).

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 (DF-GLS) unit root test of Elliott *et al.* (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.

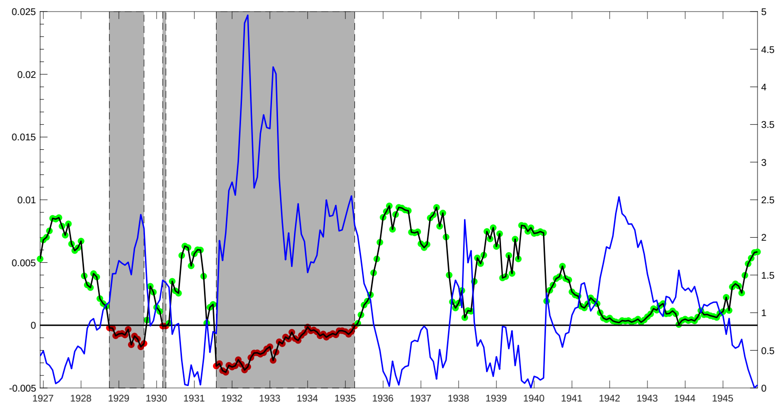
CY show that the Bonferroni Q test procedure has well controlled size and good power properties regardless of the value of the non-centrality parameter c and the degree of endogeneity of the predictor. These excellent empirical properties are, however, predicated on the key assumption that the initial condition of the predictor series x_t, defined as the deviation of the initial value of the series from its underlying mean, is asymptotically negligible. However, given that the predictor series is often well modelled by a strongly persistent near-integrated autoregressive process with $\rho = 1 - c/T$, $c > 0$, it is arguably more natural to assume that the initial condition has the same order of magnitude as the remainder of the x_t series, i.e. $O_p(T^{1/2})$, in which case the initial condition is asymptotically non-negligible and can influence the large sample properties of the Bonferroni Q test procedure. Exploring this issue forms the main focus of our paper, and we begin by briefly outlining why the size of initial condition might be expected to have a substantial impact on both the size and power of the Q test, demonstrating this with a motivating empirical example.

The Bonferroni Q procedure of CY relies on use of the DF-GLS statistic to construct a confidence interval for ρ. Müller and Elliott (2003) show that the power of the DF-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 DF-GLS test has excellent power properties when ρ is near-integrated. However, where the initial condition is asymptotically non-negligible, they show that the local alternative distribution of the DF-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 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, similar to one of those performed by CY, 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 1926M12-1994M12 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 $t = t_s, ..., T$ across multiple start dates $t_s = 1, ..., T-49$, giving a minimum sample size of 50 observations (the minimum sample size considered by CY in their Monte Carlo simulations). The results of this exercise are summarised 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 signalling a rejection (green highlights) and a lower bound below zero signalling non-rejection (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 Equation (20) below), 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

Figure 1: Lower Bound of Confidence Interval of Bonferroni Q test and $|\hat{\theta}|$



Lower Bound of CI: —— (Left Axis), $|\hat{\theta}|$: —— (Right Axis)

a majority 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 paper 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 completely robust to an asymptotically non-negligible 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 non-negligible. This is, in part, because the Bonferroni t-test of CES bases its confidence interval for ρ on the OLS demeaned Dickey-Fuller unit root statistic (DF-OLS), rather than the DF-GLS statistic, and it is known from Müller and Elliott (2003) that DF-OLS is considerably more robust than DF-GLS to the value of the initial condition.

Our first proposed testing procedure is a union-of-rejections strategy in which the null of no predictability is rejected if either the Bonferroni Q test or the Bonferroni t-test rejects. Our second proposed testing procedure 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 test and Bonferroni t-test, calibrated such that greater weight is placed on the Bonferroni Q test when the estimated magnitude of the initial condition is small, while greater weight is placed on the Bonferroni t-test when the estimated magnitude of the initial condition is large. We will show that our proposed hybrid 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 the case of 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 DF-GLS and DF-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 similar switching strategy to those developed by Elliott *et al.* (2015) and Harvey *et al.* (2021), whereby the conventional regression t-test with standard normal critical values is implemented 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 remainder of the paper is organised as follows. The predictive regression model and assumptions are detailed in section 2. An outline of the extant Bonferroni Q test of CY and Bonferroni t-test of CES is given in section 3, along with two variants of these procedures we introduce. In section 4, the asymptotic behaviour of these four tests is examined when the initial condition is asymptotically non-negligible. Our proposed hybrid test procedures are outlined in section 5, and their local asymptotic power is compared with that of the extant tests of CY and CES. In section 6 we report results of an empirical application of our proposed test procedures to the dataset utilised by CY. An on-line supplementary appendix provides a more extensive set of Monte Carlo simulations examining the local asymptotic power of our proposed tests in a wider range of scenarios, as well as a large set of simulations exploring their finite sample properties. This supplement also contains a detailed breakdown of the results in the empirical application for each returns/predictor pairing considered by CY and proofs of our main technical results. In what follows, $\overset{p}{\to}$ and $\overset{w}{\to}$ denote convergence in probability and weak convergence, respectively, as the sample size, T, diverges to infinity; $\lfloor . \rfloor$ denotes the integer part of its argument; $I(.)$ denotes the indicator function that takes a value of 1 when its argument is true, 0 otherwise; and $x := y$ ($x =: y$) indicates that x is defined by y (y is defined by x).

2 The Predictive Regression Model and Assumptions

We consider the following predictive regression model

$$r_t = \alpha + \beta x_{t-1} + u_t, \qquad t = 1, ..., T \tag{1}$$

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

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

$$w_t = \rho w_{t-1} + v_t, \qquad t = 1, ..., T \tag{3}$$

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* $\omega_v^2 := \lim_{T\to\infty} E(\sum_{t=1}^{T} v_t)^2 = \sigma_e^2 \psi(1)^2$ *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 pages 56-57 of CY). The assumptions placed on z_t allow the sequence of innovations to be conditionally heteroskedastic but imposes unconditional homoskedasticity. 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 formalised in Assumptions S.1, S.2, S.3 and W below. $\diamondsuit$

As discussed in section 1, our focus in this paper 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, $H_0 : \beta = 0$ in (1).[2] Our focus is on developing tests that offer reliable levels of size and power under different assumptions regarding the degree of persistence in the predictor variable x_t, and also 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.*

[2]The methods which we outline in this paper 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 $\beta_0 = 0$.

Assumption S.2. *The predictor $\{x_t\}$ is strongly persistent, with the autoregressive parameter ρ in (3) given by $\rho = 1 - c/T$ with c a finite non-zero 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 the autoregressive parameter 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 $\theta \sim N(\mu_\theta I(\sigma_\theta^2 = 0), \sigma_\theta^2)$. 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 autoregressive 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 paper 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 $O_p(T^{1/2})$, i.e. the initial condition is asymptotically non-negligible. 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 initialise an explosive predictor at an asymptotically negligible initial value, because otherwise the behaviour of the predictor becomes dominated by the initialisation (increasingly so over time), something which is unlikely to be credible for macroeconomic and financial variables. Hence, we do not consider asymptotically non-negligible 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 $|\rho| < 1$. $\diamondsuit$

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". An argument could be made for Assumption S.3 in this case on the basis that the initial condition is then of the same order of magnitude as the rest of the $\{w_t\}$ series. As we will subsequently show, the local asymptotic powers of tests for predictability 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 behaviour of predictive regression tests under different initial condition assumptions and magnitudes in the context of near-integrated predictors.

3 The Bonferroni Q and t Tests

In this section we outline the Q test proposed by CY and the standard t-statistic and discuss why these test statistics are infeasible for testing for predictability in their raw form. We then present the Bonferroni method proposed by CY and CES, which is used to implement feasible versions of these tests when Assumption S.1 or S.2 holds, outlining how they are performed in practice.

Consider first the Q test of CY. This is based on the statistic

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

where $x_{t-1}^{\mu} := x_{t-1} - T^{-1}\sum_{s=1}^{T} x_{s-1}$ denotes the OLS demeaned (lagged) predictor, and σ_v^2 denotes the short run variance of v_t. CY show that one-sided hypothesis tests based on this statistic possess desirable optimality properties under certain conditions (including the assumption of Gaussianity); see CY pp.31–32 for further details. Although $Q(c)$ has a standard normal limiting null distribution, tests based on $Q(c)$ are infeasible. In particular, while the parameters σ_e, σ_v, σ_{ue}, ω_v and δ are all consistently estimable and, hence, can be replaced in the expression for $Q(c)$ above by these estimates, computation of $Q(c)$ also requires knowledge of ρ, thereby implicitly requiring knowledge of the unknown nuisance parameter c, which cannot be consistently estimated. However, the Q statistic can be constructed for a specified value of c, which we denote by $\tilde{c}$, and we denote the resulting statistic $Q(\tilde{c})$. As outlined below, CY then suggest obtaining a feasible confidence interval for β based on $Q(\tilde{c})$ with the confidence interval for c obtained from inversion of the DF-GLS unit root test used to determine the values of $\tilde{c}$.

Next consider the standard t-statistic for testing the null hypothesis of $\beta = 0$, based on OLS estimation of (1), which we denote by t. The limiting distribution of the t-statistic in the local-to-unity setting is a function of the unknown non-centrality parameter c; however, a confidence interval for β based on the t test can again be constructed by making use of a confidence interval for c obtained from inversion of a unit root test. CES propose such a procedure based on using the DF-OLS unit root test, as outlined below.

Given that CY and CES derive their tests under the assumption of an $o_p(T^{1/2})$ initial condition, we first reproduce their $Q(\tilde{c})$ and t limiting distributions when Assumption S.1 or S.2 holds, and the local-to-zero alternative is given by $H_b : \beta = T^{-1}b$, where b is a finite constant (the null hypothesis, $H_0 : \beta = 0$, obtains on setting $b = 0$). The limits are given in the following Theorem (see CY and CES for the proofs).

Theorem 1. *Let data be generated according to (1)-(3). Let $(W_u(s), W_e(s))$ be a two-dimensional Weiner process with correlation parameter δ, and let $W_{e,c}(s)$ be the Ornstein-Uhlenbeck process defined by the stochastic differential equation $dW_{e,c}(s) = cW_{e,c}(s)ds + dW_e(s)$ with initial condition $W_{e,c}(0) = 0$. If Assumption S.1 or S.2 holds, then under the local alternative $H_b : \beta = T^{-1}b$,*

$$(a) \quad t \xrightarrow{w} \frac{b\omega_v \kappa_c}{\sigma_u} + \delta \frac{\tau_c}{\kappa_c} + (1-\delta^2)^{1/2} Z \tag{4}$$

$$(b) \quad Q(\tilde{c}) \xrightarrow{w} \frac{b\omega_v \kappa_c}{\sigma_u (1-\delta^2)^{1/2}} + \frac{\delta(\tilde{c}-c)\kappa_c}{(1-\delta^2)^{1/2}} + Z \tag{5}$$

where $\kappa_c := (\int_0^1 W^{\mu}_{e,c}(s)^2 ds)^{1/2}$ and $\tau_c := \int_0^1 W^{\mu}_{e,c}(s) dW_e(s)$ with $W^{\mu}_{e,c}(s) := W_{e,c}(s) - \int_0^1 W_{e,c}(r)dr$, and where Z is a standard normal random variable independent of $W_e(s)$.

Both CY and CES use Bonferroni-based methods to implement test procedures based on these statistics for an unknown value of c. Specifically, to construct a Bonferroni confidence interval for β one first constructs a $100(1-\alpha_1)\%$ confidence interval for ρ, denoted $C_\rho(\alpha_1)$. Following this, a $100(1-\alpha_2)\%$ confidence interval for β given ρ, denoted $C_{\beta|\rho}(\alpha_2)$, is calculated for each value of ρ in $C_\rho(\alpha_1)$. An overall confidence interval for β that does not depend on ρ can then be constructed as $C_\beta(\alpha) = \cup_{\rho \in C_\rho(\alpha_1)} C_{\beta|\rho}(\alpha_2)$ which, by Bonferroni's inequality, has coverage of at least $100(1-\alpha)\%$ where $\alpha := \alpha_1 + \alpha_2$. For their Bonferroni Q test, CY suggest using the DF-GLS unit root statistic to construct the initial confidence interval for ρ, while CES adopt the DF-OLS statistic when constructing the initial confidence interval for their Bonferroni t test. We now outline how these Bonferroni tests are performed.

3.1 The Bonferroni Q Test

Following the online appendix to CY,[3] performing the Bonferroni Q test proceeds as follows. First, regression (1) is estimated by OLS to obtain $\hat{\beta}$, the usual error variance estimator $\hat{\sigma}_u^2$ using the residuals $\hat{u}_t$, and the standard error for $\hat{\beta}$, denoted $SE(\hat{\beta})$. Second, the ADF regression

$$\Delta x_t = \pi + \phi x_{t-1} + \sum_{i=1}^{p-1} \psi_i \Delta x_{t-i} + e_t \tag{6}$$

is estimated by OLS, and the error variance estimator $\hat{\sigma}_e^2$ is obtained from the residuals $\hat{e}_t$, together with the long run variance estimator $\hat{\omega}_v^2 := \hat{\sigma}_e^2/(1-\sum_{i=1}^{p-1}\hat{\psi}_i)^2$. Using the estimate of the covariance given in CY, $\hat{\sigma}_{ue} := (T-p-1)^{-1}\sum_{t=p}^{T}\hat{u}_t\hat{e}_t$, the correlation parameter estimator $\hat{\delta} := \hat{\sigma}_{ue}/\hat{\sigma}_u\hat{\sigma}_e$ is then calculated. Third, OLS is used to estimate

$$x_t = \pi + \rho x_{t-1} + v_t \tag{7}$$

and obtain $\hat{\rho}$, the short run variance estimator $\hat{\sigma}_v^2$ using the residuals $\hat{v}_t$, and the standard error for $\hat{\rho}$, denoted $SE(\hat{\rho})$. Fourth, the DF-GLS unit root statistic of Elliott *et al.* (1996) is obtained as the t-statistic for testing $\zeta_0 = 0$ in the following regression estimated by OLS

$$\Delta \breve{x}_t = \zeta_0 \breve{x}_{t-1} + \sum_{i=1}^{p-1} \zeta_i \Delta \breve{x}_{t-i} + \breve{e}_t \tag{8}$$

where $\breve{x}_t := x_t - \breve{\mu}$, and $\breve{\mu}$ is obtained from the OLS regression of $x_{\breve{c}} := [x_0, (1-\breve{\rho}L)x_1, ..., (1-\breve{\rho}L)x_T]$ onto $[1, (1-\breve{\rho}), ..., (1-\breve{\rho})]$ where $\breve{\rho} := 1 - \breve{c}/T$ with $\breve{c} = 7$. In the next step, this DF-GLS test is inverted to produce a $100(1-\alpha_1)\%$ (asymptotic) confidence interval

[3] https://scholar.harvard.edu/campbell/publications/implementing-econometric-methods-efficient-tests-stock-return-predictability-0

for the non-centrality parameter c, using pre-computed confidence belts;[4] this confidence interval is denoted $[\underline{c}, \bar{c}]$ and the corresponding confidence interval for ρ is given by $[\underline{\rho}, \bar{\rho}] = [1-\bar{c}/T, 1-\underline{c}/T]$. Finally, for each of the two values $\rho = \{\underline{\rho}, \bar{\rho}\}$ an equal tailed $100(1-\alpha_2)\%$ confidence interval for β given ρ is obtained by running regression (1), but with r_t replaced by $r_t - \hat{\sigma}_{ue}(\hat{\sigma}_e\hat{\omega}_v)^{-1}(x_t - \underline{\rho}x_{t-1})$ and $r_t - \hat{\sigma}_{ue}(\hat{\sigma}_e\hat{\omega}_v)^{-1}(x_t - \bar{\rho}x_{t-1})$, respectively. If we denote the estimated coefficient on x_{t-1} in these regressions as $\hat{\beta}(\underline{\rho})$ and $\hat{\beta}(\bar{\rho})$, respectively, then a $100(1-\alpha_2)\%$ confidence interval for β is given by $[\underline{\beta}^Q(\bar{\rho}, \alpha_2), \bar{\beta}^Q(\underline{\rho}, \alpha_2)]$ where

$$\begin{aligned} \underline{\beta}^Q(\bar{\rho}, \alpha_2) &:= \hat{\beta}(\bar{\rho}) + \frac{T-2}{2}\frac{\hat{\sigma}_{ue}}{\hat{\sigma}_e\hat{\omega}_v}\left(\frac{\hat{\omega}_v^2}{\hat{\sigma}_v^2} - 1\right) SE(\hat{\rho})^2 - z_{\alpha_2/2}(1-\hat{\delta}^2)^{1/2} SE(\hat{\beta}) \qquad (9) \\ \bar{\beta}^Q(\underline{\rho}, \alpha_2) &:= \hat{\beta}(\underline{\rho}) + \frac{T-2}{2}\frac{\hat{\sigma}_{ue}}{\hat{\sigma}_e\hat{\omega}_v}\left(\frac{\hat{\omega}_v^2}{\hat{\sigma}_v^2} - 1\right) SE(\hat{\rho})^2 + z_{\alpha_2/2}(1-\hat{\delta}^2)^{1/2} SE(\hat{\beta}) \end{aligned}$$

and where $z_{\alpha_2/2}$ denotes the $\alpha_2/2$ quantile of the standard normal distribution. This confidence interval for β has (asymptotic) coverage of *at least* $100(1-\alpha)\%$, where it is recalled that $\alpha = \alpha_1 + \alpha_2$, rather than exactly $100(1-\alpha)\%$. For a right tailed test, the null of no predictability is rejected if $\underline{\beta}^Q(\bar{\rho}, \alpha_2) > 0$, whereas for a left tailed test the null is rejected if $\bar{\beta}^Q(\underline{\rho}, \alpha_2) < 0$. For a given value of δ, one-sided tests for predictability based on this confidence interval will have asymptotic size that is at most $100(\alpha/2)\%$ across all values of c. CY show, however, that tests based on this confidence interval are very conservative in practice, with the asymptotic size of one-sided tests well below $100(\alpha/2)\%$ for all values of c. As a result, CY propose a refinement that shrinks the confidence interval for ρ such that a test for β with a given (asymptotic) significance level is achieved. Denoting the desired (asymptotic) significance level for the confidence interval for β as $\tilde{\alpha}$ then, for a given value of δ, this is achieved by fixing the value of α_2 and numerically searching to find values of $\bar{\alpha}_1^Q$ and $\underline{\alpha}_1^Q$ such that

$$Pr(\underline{\beta}^Q(\bar{\rho}(\bar{\alpha}_1^Q), \alpha_2) > \beta) \leq \tilde{\alpha}/2 \text{ and } Pr(\bar{\beta}^Q(\underline{\rho}(\underline{\alpha}_1^Q), \alpha_2) < \beta) \leq \tilde{\alpha}/2 \qquad (10)$$

hold across a grid of values of $c \in [-5, 50]$. For a given value of δ the one-sided tests for predictability constructed in this manner will have an asymptotic size of exactly $\tilde{\alpha}/2$ for some value of c while remaining slightly undersized for all other values of c. Consequently,

[4]These confidence belts, along with the corresponding confidence belts based on the DF-OLS test which will be used in section 3.2 below, are available from Motohiro Yogo's personal website: https://sites.google.com/site/motohiroyogo/research/asset-pricing

two-sided tests will have size of at most $\tilde{\alpha}$ for all values of c. CY calibrate this procedure by fixing $\tilde{\alpha} = \alpha_2 = 0.1$ such that their resultant one-sided tests have a maximum (asymptotic) size of 5%. The appropriate values of $\underline{\alpha}_1^Q$ and $\bar{\alpha}_1^Q$ are reported in Table 2 of CY and are reproduced in Table 1 of this paper for ease of reference. The confidence interval for β from this size-controlled Bonferroni Q test constructed using the DF-GLS test is denoted by $[\underline{\beta}^{Q^{GLS}}(\bar{\rho}(\bar{\alpha}_1^Q), \alpha_2), \bar{\beta}^{Q^{GLS}}(\underline{\rho}(\underline{\alpha}_1^Q), \alpha_2)]$ and we denote the predictability test based on this confidence interval as Q^{GLS}. In the remainder of this paper we will follow CY and calibrate our proposed test procedures such that they deliver one-sided tests for predictability with asymptotic size of at most 5%. In practice δ is unknown, and so one proceeds as above replacing δ by its consistent estimate $\hat{\delta}$.

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 δ is estimated to be positive 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. $\diamondsuit$

3.2 The Bonferroni t Test

We now outline how the Bonferroni t test originally proposed by CES is performed. First, the DF-OLS statistic from (6) is calculated and a $100(1-\alpha_1)\%$ (asymptotic) confidence interval for c is computed using pre-computed confidence belts for the DF-OLS test. We again denote this confidence interval $[\underline{c}, \bar{c}]$ and the associated confidence interval for ρ by $[\underline{\rho}, \bar{\rho}] = [1 - \bar{c}/T, 1 - \underline{c}/T]$.

Denote by $d_{c,\eta}$ the η-level critical value of the null distribution of t for a given value of c (i.e. the critical value obtained from (4) with $b = 0$). CES show that a $100(1-\alpha_2)\%$ asymptotic confidence interval for β can be constructed as $[\underline{\beta}^t(\alpha_1, \alpha_2), \bar{\beta}^t(\alpha_1, \alpha_2)]$ where

$$\begin{aligned} \underline{\beta}^t(\alpha_1, \alpha_2) &:= \hat{\beta} - \bar{d}(\alpha_1, \alpha_2) SE(\hat{\beta}) \\ \bar{\beta}^t(\alpha_1, \alpha_2) &:= \hat{\beta} - \underline{d}(\alpha_1, \alpha_2) SE(\hat{\beta}) \end{aligned} \tag{11}$$

with

$$(\underline{d}(\alpha_1, \alpha_2), \bar{d}(\alpha_1, \alpha_2)) = \left(\min_{\underline{c} \le c \le \bar{c}} d_{c,\alpha_2/2}, \max_{\underline{c} \le c \le \bar{c}} d_{c,1-\alpha_2/2} \right).$$

Similar to the Bonferroni Q test, the null of no predictability is rejected when $\underline{\beta}^t(\alpha_1, \alpha_2) > 0$ for right tailed tests and when $\bar{\beta}^t(\alpha_1, \alpha_2) < 0$ for left tailed tests. This confidence interval for β has asymptotic coverage of at least $100(1-\alpha)\%$, where again $\alpha = \alpha_1 + \alpha_2$. However, as with the Bonferroni confidence interval obtained for the Q test above, this confidence interval for β can be conservative with a coverage rate in excess of $100(1-\alpha)\%$. In order to achieve a test for β with asymptotic size maximised at $\tilde{\alpha}$, one can fix the value of α_2 and numerically search for values of $\bar{\alpha}_1^t$ and $\underline{\alpha}_1^t$ such that

$$Pr(\underline{\beta}^t(\bar{\alpha}_1^t, \alpha_2) > \beta) \leq \tilde{\alpha}/2 \text{ and } \quad Pr(\bar{\beta}^t(\underline{\alpha}_1^t, \alpha_2) < \beta) \leq \tilde{\alpha}/2$$

holds. If $\tilde{\alpha} = \alpha_2$ is fixed at 0.1, such that the resultant one-sided tests have a maximum asymptotic size of 5% for $c \in [-5, 50]$, the appropriate values of $\underline{\alpha}_1^t$ and $\bar{\alpha}_1^t$ are those of CY, which are reported in Table 1 for convenience. The confidence interval for β from this size-controlled Bonferroni t-test constructed using the DF-OLS test is denoted by $[\underline{\beta}^{t^{OLS}}(\bar{\alpha}_1^t, \alpha_2), \bar{\beta}^{t^{OLS}}(\underline{\alpha}_1^t, \alpha_2)]$ and we denote the predictability test based on this confidence interval as t^{OLS}.

3.3 Variant Bonferroni Q and t Tests

As discussed in section 1, it is well documented in the context of testing for a unit root that the power profiles of the DF-OLS and DF-GLS unit root tests, and hence the confidence intervals they deliver for ρ, vary depending on the magnitude of the initial condition; see, in particular, Müller and Elliott (2003). Consequently, it is to be expected that the initial condition will have an impact on the performance of the Bonferroni Q and t tests through their use of the DF-GLS and DF-OLS statistics, respectively, to create confidence intervals for ρ. As we will see later, the limiting distributions of the $Q(\tilde{c})$ and t statistics also depend on the magnitude of the initial condition in the strongly persistent case. While CY and CES, respectively, propose use of DF-GLS in connection with Q and DF-OLS in connection with t, when obtaining the confidence interval for c, it is equally possible to consider using DF-OLS in connection with Q and DF-GLS in connection with t. In what follows, we consider these two variants, in addition to the originally proposed versions, and investigate the impact of the initial condition magnitude on all four of these procedures. We will refer to the Bonferroni Q test that uses the DF-OLS test as Q^{OLS}, with associated confidence interval $[\underline{\beta}^{Q^{OLS}}(\bar{\rho}(\bar{\alpha}_1^Q), \alpha_2), \bar{\beta}^{Q^{OLS}}(\underline{\rho}(\underline{\alpha}_1^Q), \alpha_2)]$, and the Bonferroni t test that uses the DF-GLS test as t^{GLS} with associated confidence interval $[\underline{\beta}^{t^{GLS}}(\bar{\alpha}_1^t, \alpha_2), \bar{\beta}^{t^{GLS}}(\underline{\alpha}_1^t, \alpha_2)]$.

The appropriate values of $\underline{\alpha}_1^Q$ and $\bar{\alpha}_1^Q$ that lead to one-sided Bonferroni Q tests with maximum asymptotic size of 5% for $c \in [-5, 50]$ when using the DF-OLS test to obtain the confidence interval for ρ are provided in Table 1. Similarly, the values of $\underline{\alpha}_1^t$ and $\bar{\alpha}_1^t$ that lead to one-sided Bonferroni t tests with maximum asymptotic size of 5% for $c \in [-5, 50]$ when using the DF-GLS test to obtain the confidence interval for ρ are also provided in Table 1. These are obtained using the same methodologies as in CY and CES, using the limiting null distributions of the $Q(\tilde{c})$ and t statistics given in Theorem 1 (on setting $b = 0$), together with the usual limit distributions of DF-OLS and DF-GLS (given in, for example, Elliott *et al.*, 1996), all under Assumptions S.1 and S.2. Here and throughout the paper results were obtained by direct simulation of the limiting distributions, with the Wiener processes approximated using NIID(0,1) random variates, and with the integrals approximated by normalized sums of 1,000 steps. All simulations were performed in Gauss 8.0 using 5,000 Monte Carlo replications.

4 Behaviour of Tests when the Initial Condition is Asymptotically Non-negligible

In this section we consider the behaviour of the Q^{GLS} and t^{OLS} tests, together with the Q^{OLS} and t^{GLS} variants introduced in the previous section, when Assumption S.3 holds, i.e. 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, arguably the most natural, case where the initial condition is of the same order as the rest of the predictor series, the limit distributions of the statistics will depend on the magnitude of the initial condition. We now quantify this dependence by deriving these limit distributions and investigating the impact of the initial condition on the local asymptotic power of the different tests. We concentrate on tests for predictability when $\delta < 0$ 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 $\delta < 0$, for the reasons outlined in Remark 3.1.

In the following Theorem, we present the limit distributions of the statistics $Q(\tilde{c})$ and t under Assumption S.3, providing an analogue to the results of Theorem 1 for the case of an asymptotically non-negligible initial condition.

Theorem 2. *Let data be generated according to (1)-(3). Let $W_u(s)$, $W_e(s)$ and $W_{e,c}(s)$ be as defined in Theorem 1. If Assumption S.3 holds then under the local alternative $H_b : \beta = T^{-1}b$,*

$$(a) \quad t \xrightarrow{w} \frac{b\omega_v \kappa_c^\theta}{\sigma_u} + \delta \frac{\tau_c^\theta}{\kappa_c^\theta} + (1-\delta^2)^{1/2} Z := t^\infty \tag{12}$$

$$(b) \quad Q(\tilde{c}) \xrightarrow{w} \frac{b\omega_v \kappa_c^\theta}{\sigma_u (1-\delta^2)^{1/2}} + \frac{\delta(\tilde{c}-c)\kappa_c^\theta}{(1-\delta^2)^{1/2}} + Z \tag{13}$$

where $\kappa_c^\theta := (\int_0^1 K_c^\mu(s)^2 ds)^{1/2}$ and $\tau_c^\theta := \int_0^1 K_c^\mu(s) dW_e(s)$ with $K_c^\mu(s) := K_c(s) - \int_0^1 K_c(r)dr$ and

$$K_c(r) := \theta(e^{-rc} - 1)(2c)^{-1/2} + W_{e,c}(r) \tag{14}$$

and where $Z \sim N(0,1)$ is a standard normal random variable that is independent of $W_e(s)$.

Remark 4.1. Observe that when $\theta = 0$, $K_c(r)$ in (14) reduces to $W_{e,c}(r)$ and hence $\kappa_c^\theta = \kappa_c$ and $\tau_c^\theta = \tau_c$, with the limit distributions for $Q(\tilde{c})$ and t given in Theorem 2 under Assumption S.3 simplifying to the limits given in Theorem 1 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, note that the limits of the tests under the $c = 0$ case of Assumption S.1, as given in Theorem 1, can be obtained from Theorem 2 on replacing $K_c(r)$ in (14) with $W_{e,0}(r) = W_e(r)$. Hence, asymptotic analysis of the tests under Assumption S.1 can again be subsumed under Theorem 2, on defining $K_0(r) := W_e(r)$. ♢

Remark 4.2. Where $\theta \neq 0$ it is seen from the representations in (12) and (13) that for near-integrated predictors with $c > 0$ the asymptotic distributions of both the t and $Q(\tilde{c})$ 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 do not depend on the sign of μ_θ; that is, they are functions of the absolute value of the magnitude of the initial condition, $|\mu_\theta|$, such that the limit distributions which obtain for μ_θ and $-\mu_\theta$ coincide. ♢

Remark 4.3. Representations for the limiting null distributions of the t and $Q(\tilde{c})$ statistics obtain on setting $b = 0$ in the expressions in (12) and (13), respectively. ♢

We now evaluate the local asymptotic power of the Q^{GLS}, Q^{OLS}, t^{GLS} and t^{OLS} tests under Assumptions S.1-S.3. To do so, in addition to the limits presented in Theorem 2, we also require the limiting distributions of DF-GLS and DF-OLS, and these are given by (see, for example, Harvey *et al.*, 2009):

$$\text{DF-GLS} \xrightarrow{w} \frac{K_c(1)^2 - 1}{2\sqrt{\int_0^1 K_c(s)^2 ds}} \tag{15}$$

$$\text{DF-OLS} \xrightarrow{w} \frac{K_c^{\mu}(1)^2 - K_c^{\mu}(0)^2 - 1}{2\kappa_c^{\theta}}. \tag{16}$$

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.

For Q^{GLS} and Q^{OLS}, we first simulate draws from the limiting distributions of DF-GLS and DF-OLS using (15) and (16), respectively. 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, implemented using the values of $\bar{\alpha}_1^Q$ appropriate for δ obtained from Table 1. Recalling that when testing in the right tail that the lower bound of the confidence interval for β is obtained from the Q test performed at an $\alpha_2/2$ level of significance with $\tilde{c} = \underline{c}(\bar{\alpha}_1^Q)$, we make use of the fact that the asymptotic local power function associated with $Q(\underline{c}(\bar{\alpha}_1^Q))$ is given by $E[\Phi(h(\bar{\alpha}_1^Q, \alpha_2))]$ where $\Phi(.)$ denotes one minus the standard normal cdf and

$$h(\bar{\alpha}_1^Q, \alpha_2) := z_{\alpha_2/2} - \frac{b\omega_v\kappa_c^{\theta}}{\sigma_u(1-\delta^2)^{1/2}} - \frac{\delta(\underline{c}(\bar{\alpha}_1^Q) - c)\kappa_c^{\theta}}{(1-\delta^2)^{1/2}} \tag{17}$$

with $z_{\alpha_2/2}$ the $1-\alpha_2/2$ quantile of the standard normal distribution, e.g. 1.645 for $\alpha_2 = 0.1$. Next we simulate a draw from κ_c^{θ} and construct $h(\bar{\alpha}_1^Q, \alpha_2)$ in (17). 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^{GLS} and t^{OLS}, in each simulation replication we again first simulate a draw from the limiting distributions of DF-GLS and DF-OLS using (15) and (16), respectively, and then obtain $[\underline{c}, \bar{c}]$ using the appropriate pre-computed confidence belts, using the values of $\bar{\alpha}_1^t$ appropriate for δ obtained from Table 1. Next we simulate the limit of t using the result in Theorem 2(a), and compare this with the critical value $\bar{d}(\bar{\alpha}_1^t, \alpha_2) := \max_{\underline{c} \leq c \leq \bar{c}} d_{c,1-\alpha_2/2}$. 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 are those designed for Assumptions S.1 and S.2.

In what follows we set $\sigma_u = \omega_v = 1$ 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 non-negligible initial conditions of increasing magnitude ($\mu_\theta = 1$ and $\mu_\theta = 3$). Results for random initial conditions, available from the authors on request, were found to be qualitatively similar and hence are not reported. We consider the local-to-unity values $c = \{0, 2, 5, 20\}$[5], and local power curves are 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 the initial condition and so only one set of power results is required. All tests are performed as one-sided (asymptotic) 5% tests.

4.1 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. When $c = 0$, the results in panel (a) show that no one test's power profile dominates all others. 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 all of the tests are asymptotically size-controlled (as expected) and that the best overall local power performance is displayed by the Q^{GLS} test. The local power of this test offers substantial power gains relative to the other three tests for $c = 2, 5$, and only ever falls very slightly below that of the t^{GLS} and t^{OLS} tests for small values of b when $c = 20$. The next best performing test is arguably the t^{GLS} test, followed closely by the t^{OLS} test, with the Q^{OLS} test displaying fairly poor local power performance for larger values of c. The additional results for $c = \{10, 50\}$ reported in the supplement (see Figure S.1) show that Q^{GLS} is again arguably the best procedure, unless $c = 50$ where it lacks power relative to the t^{GLS} and t^{OLS} tests.

[5]Results for the additional cases $c = \{10, 50\}$, and for $\delta = -0.75$, are reported in the supplementary appendix.

We next turn our attention to the case where the initial condition of the predictor is asymptotically non-negligible, with $\mu_\theta = 1$. We see from the results in panels (c), (f) and (i) of Figure 2 that all of the tests remain asymptotically size-controlled for an initial condition of this magnitude, but we observe a reduction in local power for the Q^{GLS} test relative to the case of $\mu_\theta = 0$, with this effect more pronounced the greater is the value of c; indeed, for $c = 20$, Q^{GLS} becomes the worst performing test. A similar, but slightly less pronounced, reduction in asymptotic local power is also observed for the t^{GLS} test and, of the two remaining tests, the best local power performance across all values of c considered is that associated with the t^{OLS} test, with the power of this test falling only slightly below that of the Q^{GLS} test for $c = 2, 5$, but greatly exceeding it for the larger value of $c = 20$.

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

Remark 4.4. In summary, if performing right tailed tests for predictability with $\delta < 0$ when using strongly persistent data with $c > 0$, one would ideally perform the Q^{GLS} test when $\mu_\theta = 0$ holds, while for larger μ_θ, we argue that the best strategy would be to perform the t^{OLS} test. 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. ♢

4.2 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 best performing test, followed by t^{OLS}, while Q^{OLS} performs worst overall. However, when $c > 0$ with $\mu_\theta = 0$, we see from panels (b), (e) and (h) that the local power of Q^{GLS} dominates that of the other tests only for $c = 2$, with the power of this test subsequently beginning to fall below that of the t^{OLS} and t^{GLS} tests as the value of c increases. Turning to the cases where the initial condition of the predictor is asymptotically non-negligible, 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. We argue that the best overall performance is given by the t^{OLS} test, with this test displaying an attractive local power profile across the scenarios considered. While the t^{GLS} test has a noticeable power advantage over t^{OLS} when $c = 20$ and $\mu_\theta = 3$, this comes at the cost of modest oversize in this case.

Remark 4.5. 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 best overall asymptotic size and local power profile among the candidate tests 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 $\delta > 0$. ♢

5 Hybrid Tests

While it is clear from the previous section that, under strong persistence when $\delta < 0$, t^{OLS} represents the best approach to conducting left tailed tests, the situation is more complicated for right tailed testing, with Q^{GLS} being the best overall approach when $c > 0$ and $\theta = 0$ and t^{OLS} the best procedure for larger θ when $c > 0$, while there is little to choose between the two tests when $c = 0$. We now consider the right tailed testing context, and propose tests for predictability that are designed to exploit the superior performance of the Q^{GLS} and t^{OLS} tests for different initial value magnitudes when $c > 0$.

5.1 A Union-of-Rejections Strategy

Our first proposed testing procedure is a union-of-rejections strategy in which we reject the null hypothesis of $\beta = 0$ in favour 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. Such a strategy is common in the time series econometrics literature, following Harvey *et al.* (2009) in the context of unit root testing. A simple union-of-rejections test based on setting $\tilde{\alpha} = 0.1$ 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 $\tilde{\alpha}/2$ we therefore need to modify the significance levels at which the initial confidence belts for ρ are constructed for both the DF-GLS and DF-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 $\underline{\beta}^{Q^{GLS}}(\bar{\rho}(\bar{\alpha}_1^Q), \alpha_2)$ and $\underline{\beta}^{t^{OLS}}(\bar{\alpha}_1^t, \alpha_2)$, respectively, then our proposed union-of-rejections test, U, is defined by the decision rule

$$U: \text{ Reject } H_0 \text{ if } \underline{U} > 0 \tag{18}$$

where

$$\underline{U} := \max\left(\underline{\beta}^{Q^{GLS}}(\bar{\rho}(\xi\bar{\alpha}_1^Q), \alpha_2), \underline{\beta}^{t^{OLS}}(\xi\bar{\alpha}_1^t, \alpha_2)\right). \tag{19}$$

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 $\tilde{\alpha}/2$ across a specified range of values of c and initial conditions. The local limiting behaviour of $\underline{U}$ will subsequently be detailed in Theorem 3.

5.2 A Weighting Strategy

Given that the union-of-rejections test procedure outlined above is constructed in an attempt to capture the desirable properties of Q^{GLS} when $\theta = 0$ and those of t^{OLS} when θ is large, we also consider a second procedure that takes a weighted average of Q^{GLS} and t^{OLS}, with the weight being a function based around an estimate of the initial condition (relative) magnitude, θ. In the context of testing for a unit root, Harvey and Leybourne (2005) use an estimate of θ to construct a test where more weight is placed on the DF-GLS (DF-OLS) test when θ is estimated to be small (large). They propose using the following estimate of $|\theta|$,

$$|\hat{\theta}| := |x_0 - \hat{\mu}|/\hat{\sigma}_w \tag{20}$$

where $\hat{\mu} := T^{-1}\sum_{i=1}^{T} x_i$ and $\hat{\sigma}_w^2 := T^{-1}\sum_{i=1}^{T}(x_i - \hat{\mu})^2$. Under Assumption S.3, Harvey and Leybourne (2005,p.102) show that $|\hat{\theta}|$ has a well-defined limiting distribution and, hence, is not consistent for $|\theta|$. 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) 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

$$\lambda_\gamma(|\hat{\theta}|) := exp(-\gamma|\hat{\theta}|) \tag{21}$$

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

$$W_\gamma: \text{ Reject } H_0 \text{ if } \underline{W}_\gamma > 0 \tag{22}$$

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

$$\underline{W}_\gamma := \lambda_\gamma(|\hat{\theta}|)\underline{\beta}^{Q^{GLS}}(\bar{\rho}(\xi\bar{\alpha}_1^Q), \alpha_2) + (1 - \lambda_\gamma(|\hat{\theta}|))\underline{\beta}^{t^{OLS}}(\xi\bar{\alpha}_1^t, \alpha_2). \tag{23}$$

In (23), the parameter $\xi < 1$ plays the same role as it does in the context of the U test in (18), allowing us to control the maximum asymptotic size of W_γ in (22) at some desired level, $\tilde{\alpha}/2$.

In Theorem 3 we now detail the local limiting behaviour of $\underline{U}$ and $\underline{W}_\gamma$ under Assumption S.3. Recall that the corresponding limiting behaviour of $\underline{U}$ and $\underline{W}_\gamma$ under Assumptions S.1 and S.2 can be obtained by setting $c = 0$ and $\theta = 0$, respectively, in $K_c(r)$ (cf. Remark 4.1).

Theorem 3. *Let data be generated according to (1)-(3). Let $W_u(s)$, $W_e(s)$ and $W_{e,c}(s)$ be as defined in Theorem 1. If Assumption S.3 holds then under the local alternative $\beta = b/T$,*

$$\underline{U} \overset{w}{\to} max\left(\underline{\beta}^{Q,\infty}, \underline{\beta}^{t,\infty}\right) \tag{24}$$

$$\underline{W}_\gamma \overset{w}{\to} exp(-\gamma A_c^\theta)\underline{\beta}^{Q,\infty} + \left\{1 - exp(-\gamma A_c^\theta)\right\} \underline{\beta}^{t,\infty} \tag{25}$$

where $A_c^\theta := (\int_0^1 K_c^\mu(r)^2 dr)^{-1/2}(| - \int_0^1 K_c(r)dr|)$ with $K_c^\mu(s)$ as defined in Theorem 2 and

$$\underline{\beta}^{Q,\infty} := Z - h(\xi\bar{\alpha}_1^Q, \alpha_2)$$
$$\underline{\beta}^{t,\infty} := t^\infty - \bar{d}(\xi\bar{\alpha}_1^t, \alpha_2)$$

where $h(\cdot)$ and t^∞ are as defined in (17) and (12), respectively, and Z is a standard normal random variable that is independent of $W_e(s)$.

Remark 5.1. As noted above, $\underline{W}_\gamma$ 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 maximum size of all hybrid tests was always found to be within the interior of these search grids). The required ξ values, obtained by simulation of the relevant limiting distributions, are reported in Table 1.

5.3 Allowing for Weakly Persistent Predictors

When Assumption W holds, such that the predictor is weakly persistent, the Bonferroni Q and t tests discussed in section 3 are all asymptotically invalid. Moreover, in our Monte Carlo exercise reported in the supplementary appendix, we find that although the finite sample sizes of the t^{OLS} and t^{GLS} tests remain reasonably well controlled for large values of the local-to-unity parameter, c, the Q^{GLS} and Q^{OLS} tests both suffer from severe size distortions when c is large. In particular, when testing for predictability in the right and left tails with $\delta < 0$, the Q^{GLS} test can be severely oversized, whereas the Q^{OLS} test can be severely oversized for right tailed tests and severely undersized for left tailed tests. The behaviour of Q^{GLS} therefore also renders the hybrid procedures 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 therefore propose an approach, similar in spirit to that used in Elliott *et al.* (2015) and Harvey *et al.* (2021), whereby we switch from the use of the Bonferroni-based hybrid 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 *et al.* (2021), we propose using the Dickey-Fuller normalised 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 $\hat{\psi}_i$, $i = 1, ..., p-1$ are obtained by OLS estimation of (6). 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, $-\gamma_\phi T^{1/2}$, $\gamma_\phi > 0$, so that the conventional t-test is used whenever $ADF_\phi < -\gamma_\phi T^{1/2}$. The divergence rate of the ADF_ϕ statistic 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 γ_ϕ, we found that a choice of $\gamma_\phi = 4.5$ led to the best overall size control for the hybrid procedures in finite samples so will adopt this value of γ_ϕ in what follows.

5.4 Proposed Hybrid Testing Procedures

On the basis of the preceding results, we now propose 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 5.2. Our proposed decision rules for one-sided tests performed at the $\alpha/2$ nominal asymptotic level can be written as follows, where we denote the $(1 - \alpha)$ quantile of the normal distribution as z_α, and $\hat{\beta}$ and $s.e(\hat{\beta})$ the OLS estimate and standard error of β from OLS estimation of (1). All confidence intervals are constructed so that the resultant one-sided tests for predictability have maximum asymptotic size of $\alpha/2$.

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

- **Right Tailed Tests:**
  - **Decision Rule for $U^{\mathbf{hyb}}$:**
    - **If** $ADF_\phi < -4.5T^{1/2}$**: Reject** H_0 **if** $\hat{\beta} - z_{\alpha/2}s.e(\hat{\beta}) > 0$
    - **If** $ADF_\phi \geq -4.5T^{1/2}$**: Reject** H_0 **if** $\underline{U} > 0$
  - **Decision Rule for $W_\gamma^{\mathbf{hyb}}$:**
    - **If** $ADF_\phi < -4.5T^{1/2}$**: Reject** H_0 **if** $\hat{\beta} - z_{\alpha/2}s.e(\hat{\beta}) > 0$
    - **If** $ADF_\phi \geq -4.5T^{1/2}$**: Reject** H_0 **if** $\underline{W}_\gamma > 0$
- **Left Tailed Tests:**
  - **Decision Rule for $U^{\mathbf{hyb}}$ and $W_\gamma^{\mathbf{hyb}}$:**
    - **If** $ADF_\phi < -4.5T^{1/2}$**: Reject** H_0 **if** $\hat{\beta} + z_{\alpha/2}s.e(\hat{\beta}) < 0$
    - **If** $ADF_\phi \geq -4.5T^{1/2}$**: Reject** H_0 **if** $\bar{\beta}^{t^{OLS}}(\underline{\alpha}_1^t, \alpha_2) < 0$

Remark 5.2. 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 empirical practice, the true value of δ will be unknown, but the appropriate approach can be determined according to the sign of the consistent estimator, $\hat{\delta}$. $\diamondsuit$

Remark 5.3. The switching decision rule outlined above can also be applied to the original Q^{GLS} test of CY, so that one uses the Bonferroni Q^{GLS} test as outlined in CY, unless $ADF_\phi < -4.5T^{1/2}$ in which case the conventional t test is used. This switching-based testing procedure is asymptotically valid for both weakly and strongly persistent predictors, generated according to Assumptions W or S.1-S.2, respectively. It should be stressed though that this procedure is obviously not valid for strongly persistent predictors with asymptotically non-negligible initial conditions generated according to Assumption S.3. $\diamondsuit$

Remark 5.4. While the definition of the 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 asymptotic 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 asymptotic size no greater than α. ♢

5.5 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}, Q^{OLS}, t^{GLS} and t^{OLS} procedures. We report results for the same constellation of settings as in Section 4 (additional simulations exploring the case where $\delta = -0.75$, as well as a larger range of values of c, are provided 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 4.

Firstly, we note that when $c = 0$, the hybrid tests perform very well, being size controlled 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.

When $c > 0$, we first discuss the case $\mu_\theta = 0$ 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. In terms of asymptotic local power we see that, as would be 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 whose power is only marginally lower than that of 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 supplemental Figure S.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^{GLS} and t^{OLS} tests.

Next we consider the asymptotically non-negligible 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. 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 S.2 and S.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 non-negligible 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.

Monte Carlo simulation results, reported 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 proposed hybrid procedures carry over to finite sample environments. Additional simulations also show that our proposed procedures have controlled size and excellent power properties for a weakly persistent predictor, owing to them switching into the standard t-test in the limit, whereas both the Q^{GLS} and Q^{OLS} tests suffer from severe size distortions.

6 Empirical Application

We now report results of an empirical exercise in which we revisit the dataset of 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 U^{hyb} and W_γ^{hyb} procedures are able to overcome these shortcomings.[6] As a preliminary analysis we applied the Q^{GLS}, t^{OLS}, U^{hyb} and W_γ^{hyb} 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 $(y - r_1)$. All tests are performed as one-sided tests at a nominal level of 5% and the DF-OLS and DF-GLS unit root test statistics used to construct the tests for predictability were, following CY, estimated using a lag length chosen by the Bayes Information Criteria with $p_{\max} = 5$, with this lag length selection method also used for the Dickey-Fuller normalised bias coefficient unit root test statistic. Table 2 provides a summary of the results of this exercise including the full sample estimates of the correlation parameter, $\hat{\delta}$, and the Dickey-Fuller normalised bias coefficient unit root test statistic (critical value in parentheses), along with the rejection frequency of each procedure computed as the proportion of start dates for which the test rejects the null of no predictability (entries in bold highlight the procedure with the largest proportion of rejections).[7] All tests are performed as right tailed tests with the exception

[6] The full dataset used in this empirical exercise, as well as implementation code, is available from https://rtaylor-essex.droppages.com/esrc2/default.htm

[7] We do not report results for the Annual CRSP 1952-2002 example due to this sample containing only $T = 51$ observations.

of those utilising CRSP returns from 1952-2002 using r_3 as a predictor which are performed as left tailed tests due to CY finding 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 4 and 5.5. There were only 5 subsample regressions in the entire exercise for which the Dickey-Fuller normalised bias coefficient unit root test 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 1880-2002 for the S&P500 or from 1926-2002 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 1880-1994 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 1952-2002, 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 zero, with unreported simulations showing that all tests share an almost identical power profile for $\delta = 0$. Consequently, we do not consider the 1952-2002 CRSP data further.

Given that the information in Table 2 only gives us a broad overview of the behaviour of the procedures when applied across various start dates, we now focus our attention on the variability in test rejections in relation to the value of the initial condition. We concentrate on the CRSP data from 1926-1994, with results for other returns/predictor pairings explored in the supplement. Given that we will be focusing on cases where $d - p$ and $e - p$ are used as predictors, and that there is little evidence of these variables being significant predictors of returns in the post-war period based on the predictive regressions using the 1952-2002 CRSP data summarised in Table 2, we consider start dates, t_s, for the predictive regressions up to and including the end of 1945. We begin by plotting, for each predictor, the DF-OLS and DF-GLS unit root test statistics across start dates t_s, with these plots provided in Figure 5, allowing us to see the variation in the unit root test statistics used to compute the Bonferroni confidence intervals, as the start date changes. In addition, the blue line plots $|\hat{\theta}|$, the estimate of the relative magnitude of the initial condition, as defined in (20). We see that there is generally more variability in the DF-GLS statistics across start dates than for the DF-OLS statistics, and that when $|\hat{\theta}|$ increases the value of the DF-GLS statistic also increases, while the value of the DF-OLS statistic tends to decrease. These figures show that as the start date changes, the changing magnitude of the initial condition is driving movement in the behaviour of the test statistics, reflecting the fact that in the presence of a large initial condition, the DF-GLS statistic is upward biased and the DF-OLS statistic is downward biased. For these four 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 6 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 non-rejection), 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 6(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 6(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 6(d)), on the other hand, rejects for each and every start date, and the U^{hyb} (Figure 6(c)) and W_2^{hyb} (Figure 6(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 7 reports results for the quarterly CRSP returns when using $d-p$ as a predictor. For this predictor we note that all tests fail to reject the null of no predictability for most of the earliest start dates and the overall rejection frequencies for U^{hyb}, W_γ^{hyb}, $\gamma = 1, 2$, and t^{OLS} are higher overall than for Q^{GLS}. Of most interest are the start dates $t_s = 1931\text{Q}4$ through to $t_s = 1932\text{Q}4$, where U^{hyb}, W_γ^{hyb}, $\gamma = 1, 2$, and t^{OLS} reject the null while Q^{GLS} does not, with these start dates associated with some of the largest values of $|\hat{\theta}|$ for the predictor. We also note that a run of non-rejections for Q^{GLS} is observed for start dates $t_s =$ 1941Q1 through to $t_s = 1941\text{Q}4$ when $|\hat{\theta}|$ is growing in magnitude, with the U^{hyb}, W_γ^{hyb}, $\gamma = 1, 2$, and t^{OLS} procedures continuing to reject for these start dates. While the t^{OLS} test rejects with almost the same frequency as our proposed tests in this example, we note that the rejections for t^{OLS} for start dates in 1933-1937, where $|\hat{\theta}|$ is small, are considerably weaker than for our proposed tests, with the estimated lower bound of the confidence interval for β from the t^{OLS} test being far closer to zero than for our proposed tests

Figure 8 reports results for the monthly CRSP returns when using the $e-p$ predictor. 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 $t_s = 1928\text{M}10$ through to $t_s = 1929\text{M}9$ and $t_s = 1931\text{M}8$ through to $t_s = 1935\text{M}4$, 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 non-rejections associated with a period in which $|\hat{\theta}|$ is very large indeed. While the rejection frequency for the t^{OLS} test is 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 $|\hat{\theta}|$.

Finally, Figure 9 reports results for the monthly CRSP returns series with $d-p$ employed as the predictor. In this case we note that all procedures fail to reject for start dates $t_s =$ 1926M12 through to $t_s = 1931\text{M}7$, but while Q^{GLS} only begins to reject for start dates after t_s = 1933M3, the t^{OLS} test begins to reject after t_s =1931M7, the W_2^{hyb} procedure after t_s = 1931M9 and the W_1^{hyb} and U^{hyb} procedures for start dates after t_s = 1931M10, far earlier than Q^{GLS}. There are also start dates later in the sample (t_s = 1937M9 through to t_s = 1938M4 and t_s = 1941M11 through to t_s = 1942M4) for which Q^{GLS} fails to reject while U^{hyb}, W_γ^{hyb}, $\gamma = 1, 2$, and t^{OLS} continue to reject. In all instances the start dates in which U^{hyb}, W_γ^{hyb}, $\gamma = 1, 2$, and t^{OLS} reject while Q^{GLS} does not are associated with very large values of $|\hat{\theta}|$. On the other hand, for small values of $|\hat{\theta}|$ we see that the t^{OLS} test is far less likely to reject than all other tests, with the best example of this being start dates from t_s =1933M5-1936M10 for which t^{OLS} fails to reject, with all other tests rejecting fairly consistently for this long window of start dates.

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 due to the fact that the initial condition changes across start dates. 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 either the Q^{GLS} and t^{OLS} tests tallies with our asymptotic and finite sample simulation results, and reinforce our conclusion that the new hybrid procedures can deliver more consistent power across large and small magnitudes of the predictor's initial condition.

7 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 non-negligible 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 initial condition is asymptotically non-negligible. 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 analysed 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.

References

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

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

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

Elliott, G. (1999). Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. International Economic Review 40, 767-783.

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.

Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996). Efficient tests for an autoregressive unit root. Econometrica 64, 813-836.

Fama, E.F. (1981). Stock returns, real activity, inflation, and money. American Economic Review 71, 545-565.

Harvey, D.I. and Leybourne, S.J. (2005). On testing for unit roots and the initial observation. Econometrics Journal 8, 97-111.

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.

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.

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

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

Müller, U.K. and Elliott, G. (2003). Tests for unit roots and the initial condition. Econometrica 71, 1269-1286.

Phillips, P. C. B. and Magdalinos, T. (2009). Econometric inference in the vicinity of unity. Working paper, Singapore Management University.

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

Table 1: Parameters to Deliver One-Sided Tests with Maximum 5% Asymptotic Size.

	Q^{GLS}		Q^{OLS}		t^{GLS}		t^{OLS}		U	W_1	W_2
δ	$\underline{\alpha}_1^Q$	$\bar{\alpha}_1^Q$	$\underline{\alpha}_1^Q$	$\bar{\alpha}_1^Q$	$\underline{\alpha}_1^t$	$\bar{\alpha}_1^t$	$\underline{\alpha}_1^t$	$\bar{\alpha}_1^t$	ξ	ξ	ξ
-0.999	0.050	0.055	0.050	0.055	0.005	0.035	0.020	0.035	0.28	0.85	0.55
-0.975	0.055	0.080	0.055	0.065	0.005	0.035	0.025	0.035	0.28	0.85	0.56
-0.950	0.055	0.100	0.055	0.070	0.005	0.040	0.025	0.040	0.38	0.75	0.50
-0.925	0.055	0.115	0.060	0.080	0.005	0.040	0.025	0.040	0.34	0.73	0.50
-0.900	0.060	0.130	0.060	0.085	0.005	0.035	0.025	0.035	0.34	0.71	0.46
-0.875	0.060	0.140	0.060	0.090	0.005	0.035	0.025	0.035	0.32	0.72	0.45
-0.850	0.060	0.150	0.065	0.095	0.005	0.035	0.025	0.035	0.30	0.71	0.43
-0.825	0.060	0.160	0.065	0.105	0.005	0.035	0.025	0.035	0.35	0.71	0.46
-0.800	0.065	0.170	0.065	0.110	0.005	0.035	0.025	0.035	0.33	0.71	0.46
-0.775	0.065	0.180	0.065	0.115	0.015	0.050	0.030	0.035	0.36	0.72	0.56
-0.750	0.065	0.190	0.070	0.120	0.015	0.050	0.025	0.035	0.36	0.72	0.56
-0.725	0.065	0.195	0.070	0.125	0.015	0.050	0.025	0.035	0.36	0.71	0.57
-0.700	0.070	0.205	0.070	0.130	0.015	0.025	0.025	0.035	0.37	0.71	0.57
-0.675	0.070	0.215	0.075	0.135	0.015	0.020	0.025	0.035	0.34	0.71	0.57
-0.650	0.070	0.225	0.075	0.140	0.015	0.020	0.025	0.035	0.34	0.67	0.55
-0.625	0.075	0.230	0.075	0.145	0.015	0.020	0.025	0.035	0.34	0.67	0.51
-0.600	0.075	0.240	0.080	0.150	0.020	0.025	0.030	0.035	0.34	0.65	0.51
-0.575	0.075	0.250	0.080	0.155	0.035	0.025	0.035	0.035	0.32	0.65	0.45
-0.550	0.080	0.260	0.085	0.160	0.035	0.025	0.035	0.035	0.30	0.65	0.45
-0.525	0.080	0.270	0.085	0.170	0.045	0.045	0.045	0.035	0.28	0.65	0.45
-0.500	0.080	0.280	0.085	0.175	0.080	0.045	0.060	0.035	0.28	0.60	0.45
-0.475	0.085	0.285	0.090	0.180	0.100	0.045	0.050	0.035	0.20	0.58	0.36
-0.450	0.085	0.295	0.090	0.185	0.130	0.045	0.055	0.040	0.21	0.56	0.36
-0.425	0.090	0.310	0.095	0.190	0.130	0.045	0.035	0.040	0.21	0.56	0.36
-0.400	0.090	0.320	0.095	0.200	0.130	0.045	0.060	0.040	0.21	0.53	0.36
-0.375	0.095	0.330	0.100	0.205	0.130	0.020	0.040	0.040	0.18	0.45	0.36
-0.350	0.100	0.345	0.105	0.215	0.130	0.015	0.030	0.040	0.18	0.45	0.36
-0.325	0.100	0.355	0.105	0.225	0.130	0.005	0.015	0.045	0.18	0.43	0.32
-0.300	0.105	0.360	0.110	0.230	0.130	0.005	0.010	0.050	0.18	0.41	0.26
-0.275	0.110	0.370	0.115	0.235	0.130	0.005	0.005	0.040	0.18	0.41	0.25
-0.250	0.115	0.375	0.125	0.245	0.150	0.005	0.005	0.035	0.18	0.41	0.25
-0.225	0.125	0.380	0.130	0.255	0.150	0.005	0.005	0.025	0.18	0.35	0.23
-0.200	0.130	0.390	0.140	0.260	0.150	0.005	0.005	0.025	0.10	0.33	0.21
-0.175	0.140	0.395	0.145	0.270	0.150	0.005	0.005	0.010	0.10	0.33	0.21
-0.150	0.150	0.400	0.155	0.290	0.150	0.005	0.005	0.010	0.10	0.33	0.21
-0.125	0.160	0.405	0.170	0.295	0.150	0.005	0.005	0.010	0.10	0.33	0.21
-0.100	0.175	0.415	0.190	0.310	0.150	0.005	0.005	0.005	0.10	0.33	0.21
-0.075	0.190	0.420	0.205	0.330	0.150	0.005	0.005	0.005	0.10	0.33	0.21
-0.050	0.215	0.425	0.235	0.330	0.150	0.005	0.005	0.005	0.10	0.33	0.21
-0.025	0.250	0.435	0.275	0.345	0.150	0.005	0.005	0.005	0.10	0.33	0.21

Table 2: Empirical Application Summary

Series	Predictor	$\hat{\delta}$	Rejection Frequency ADF_ϕ(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%	90.9%	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%
	$y-r_1$	-0.100	-25.92 (-64.27)	**50.3%**	49.0%	**50.3%**	46.5%	46.5%
Monthly CRSP	$d-p$	-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%**

Figure 2: Asymptotic Local Power of Right Tailed Bonferroni t and Q Tests, $\delta = -0.95$

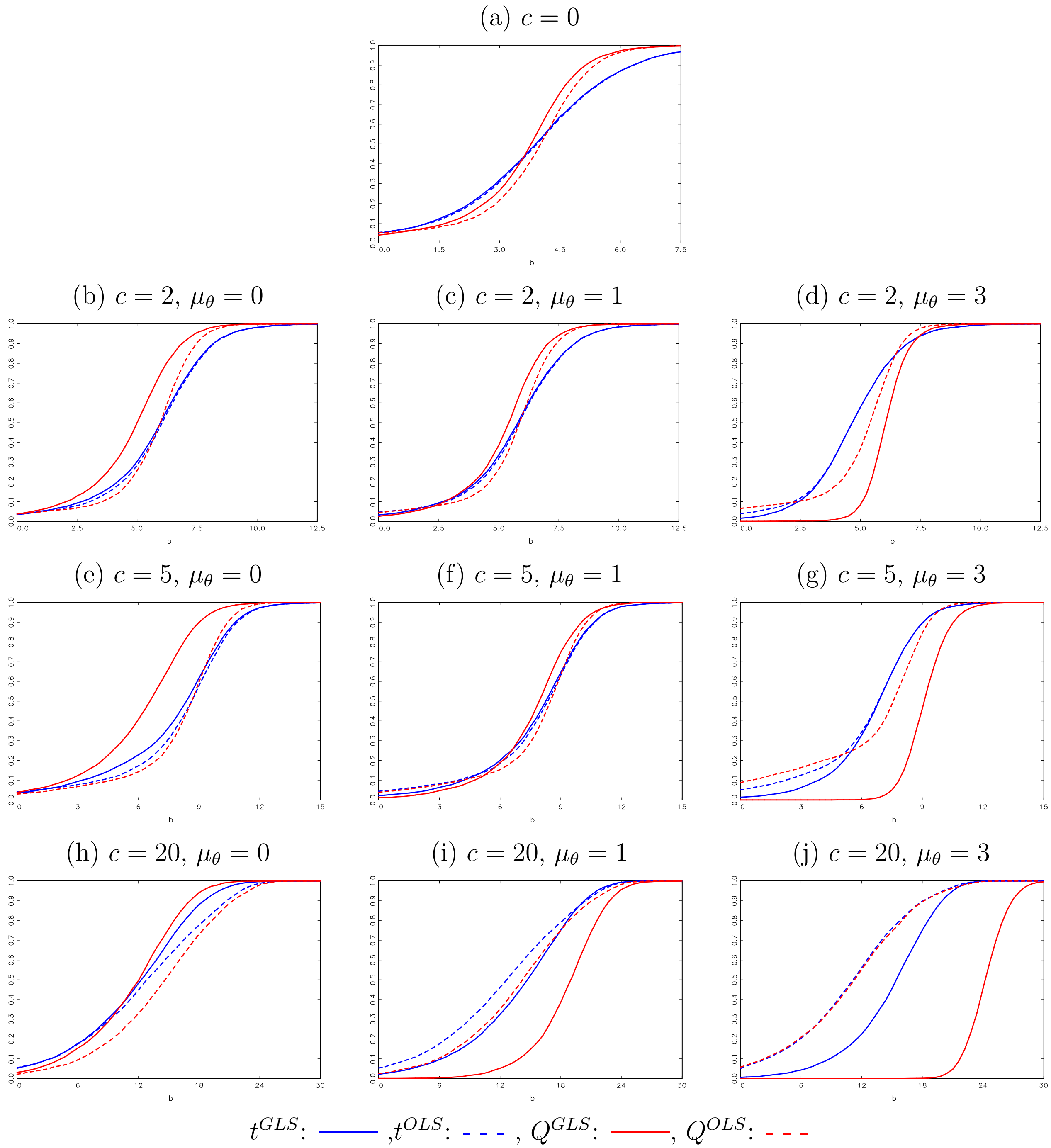


Figure 3: Asymptotic Local Power of Left Tailed Bonferroni t and Q Tests, $\delta = -0.95$

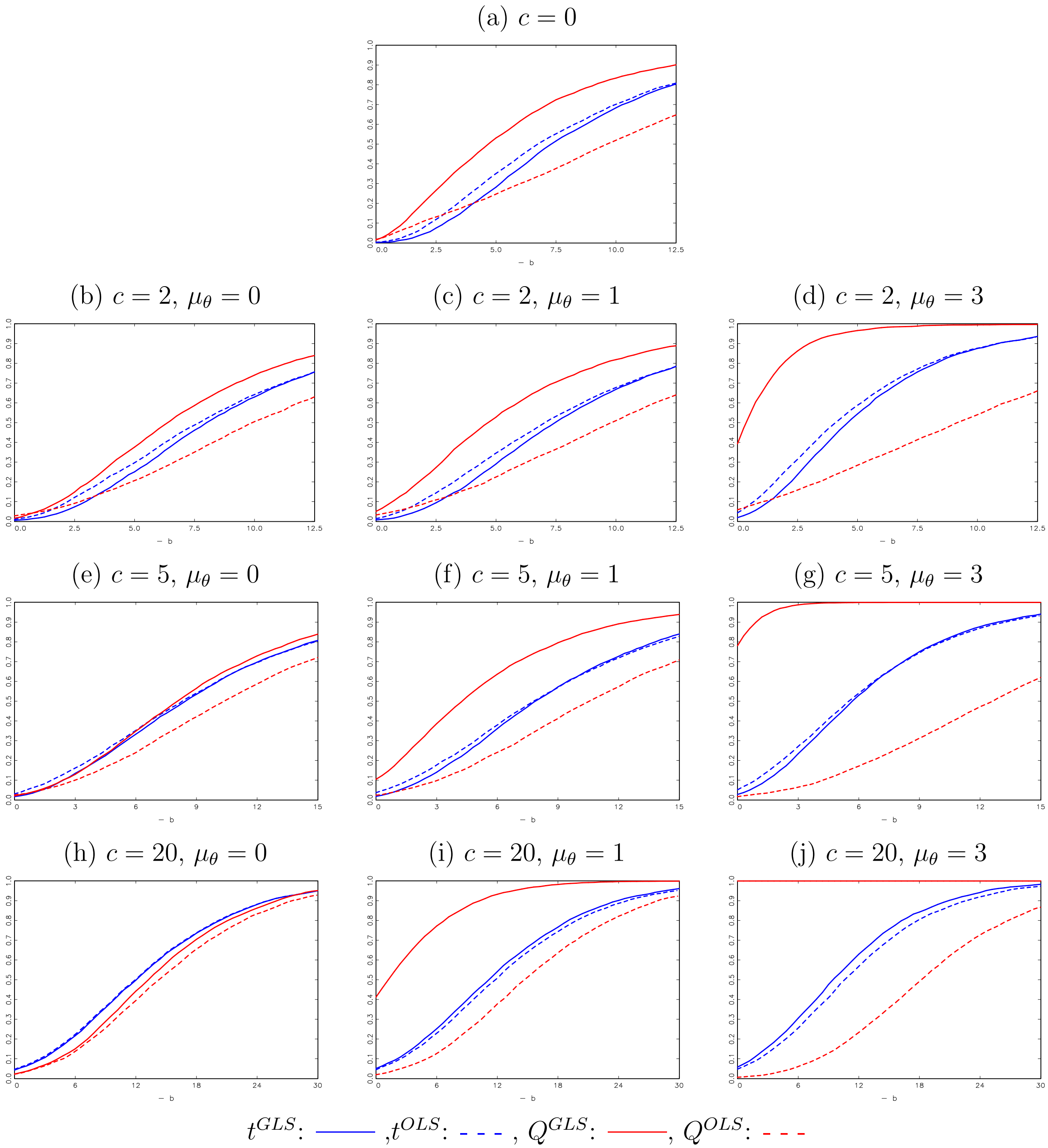
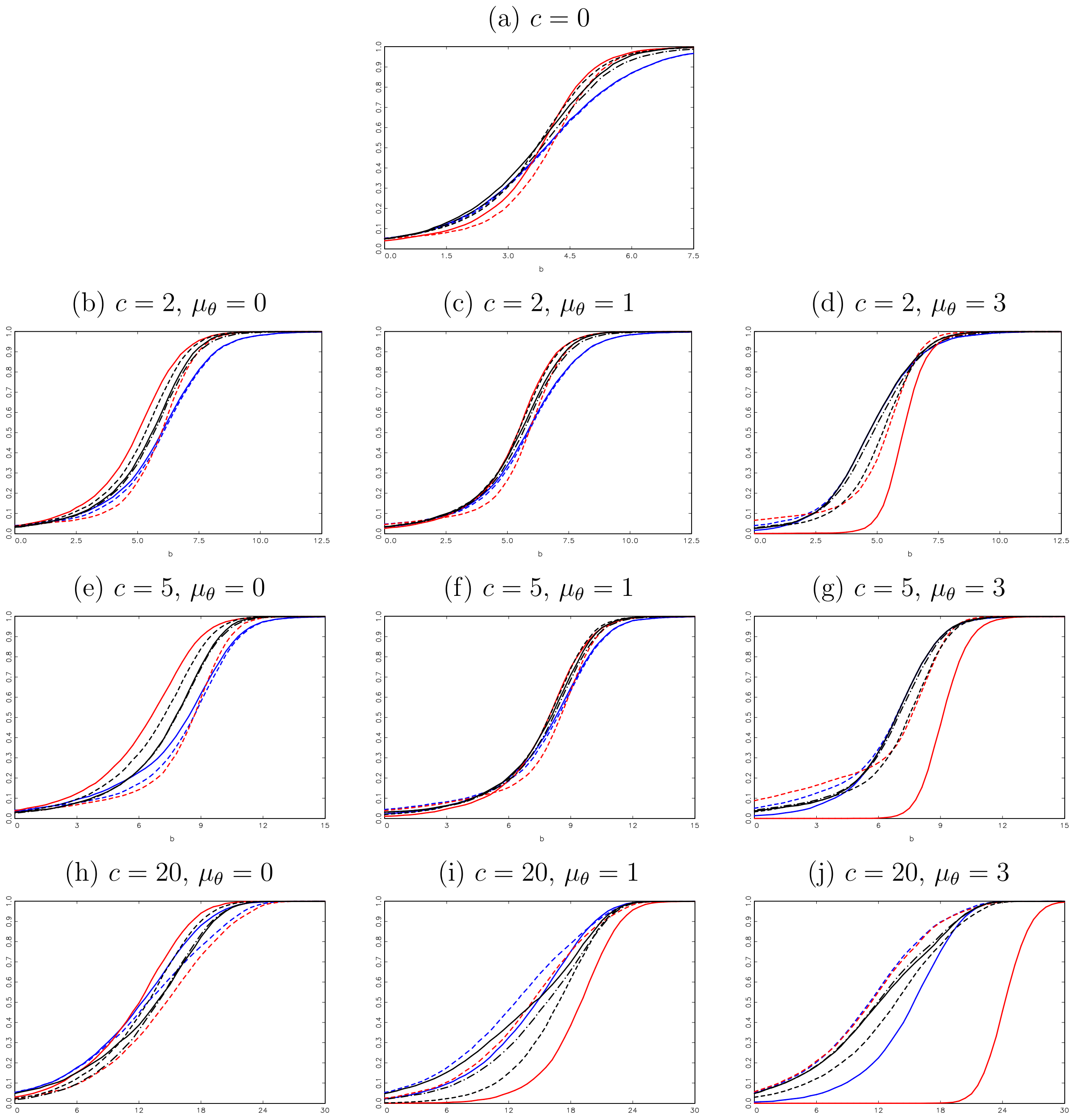


Figure 4: Asymptotic Local Power of Right Tailed Tests, $\delta = -0.95$

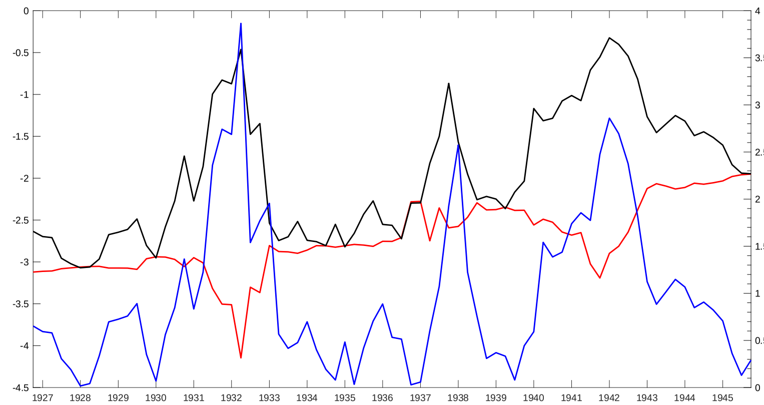


t^{GLS}: ——— ,t^{OLS}: - - - , Q^{GLS}: ———, Q^{OLS}: - - -, U^{hyb}———, W_1^{hyb}: - - -, W_2^{hyb}: —•—

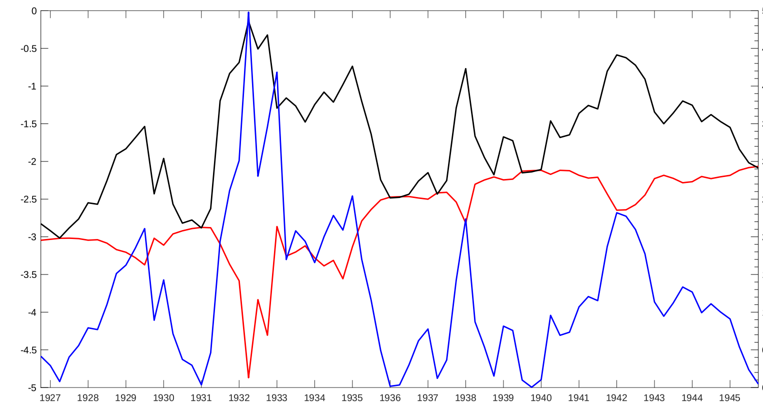
Figure 5: Unit Root Test Statistics and Estimated Magnitude of Initial Condition by Start Date - Quarterly & Monthly Data

Quarterly CRSP 1926-1994

(a) $d - p$

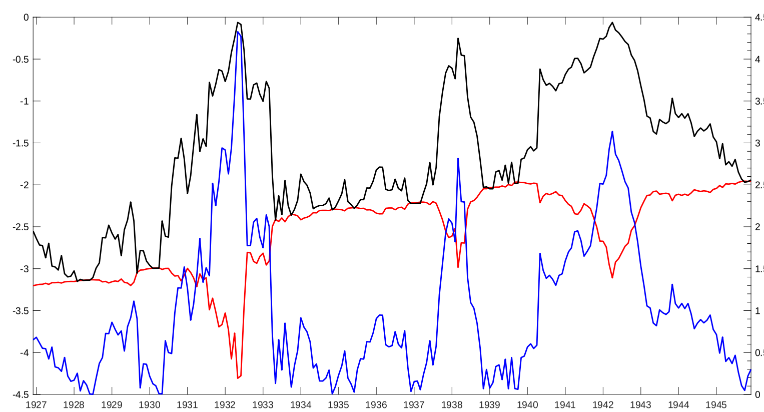


(b) $e - p$

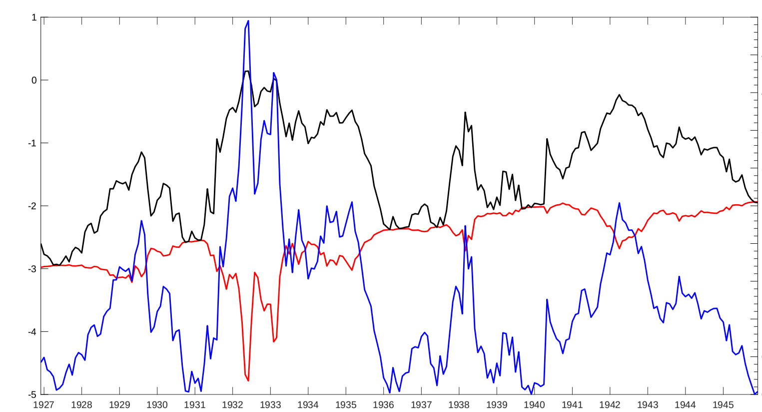


Monthly CRSP 1926-1994

(c) $d - p$



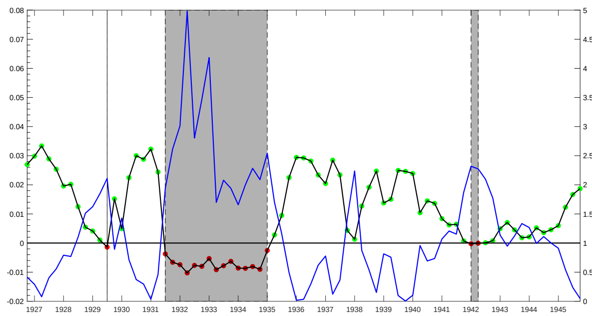
(d) $e - p$



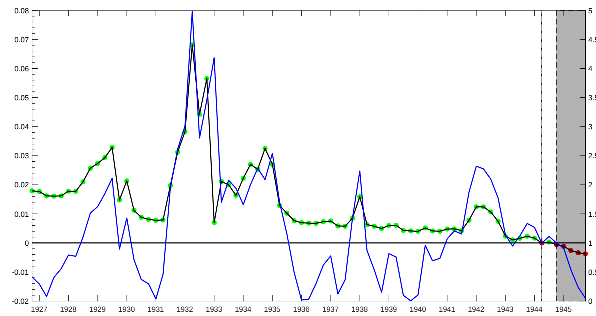
DF-OLS: —— (Left Axis), DF-GLS: —— (Left Axis), $|\hat{\theta}|$: —— (Right Axis)

Figure 6: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Quarterly CRSP 1926-1994 (Predictor = $e - p$)

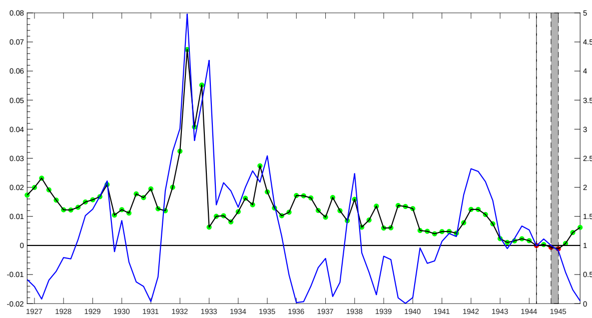
(a) Q^{GLS} (Rejection Rate 77%)



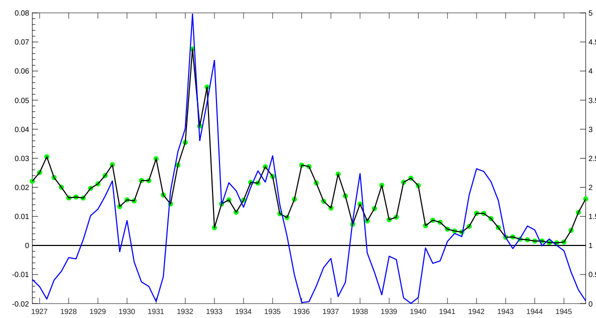
(b) t^{OLS} (Rejection Rate 92%)



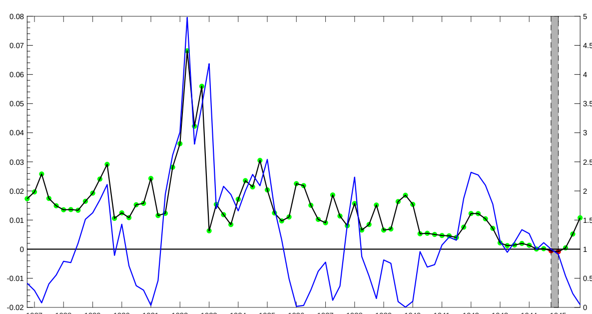
(c) U^{hyb} (Rejection Rate 96%)



(d) W_1^{hyb} (Rejection Rate 100%)



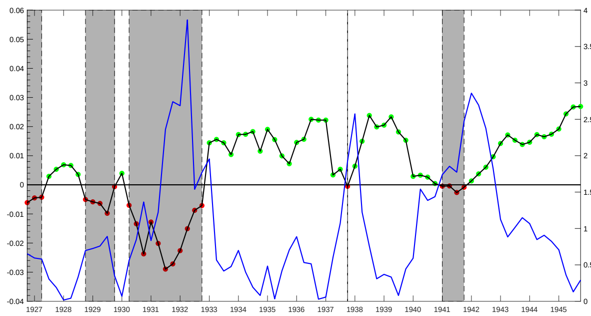
(e) W_2^{hyb} (Rejection Rate 97%)



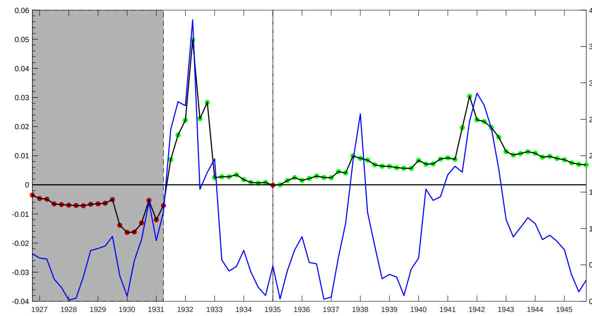
Lower Bound of CI: —— (Left Axis) ,$|\hat{\theta}|$: —— (Right Axis)

Figure 7: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Quarterly CRSP 1926-1994 (Predictor = $d - p$)

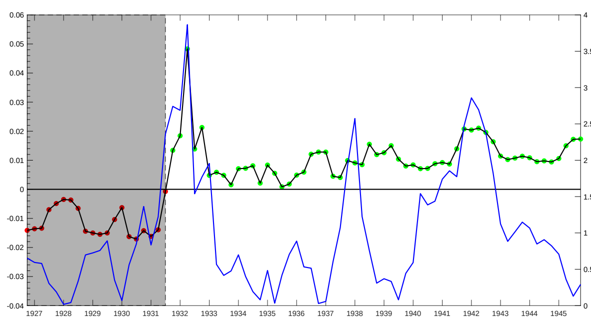
(a) Q^{GLS} (Rejection Rate 69%)



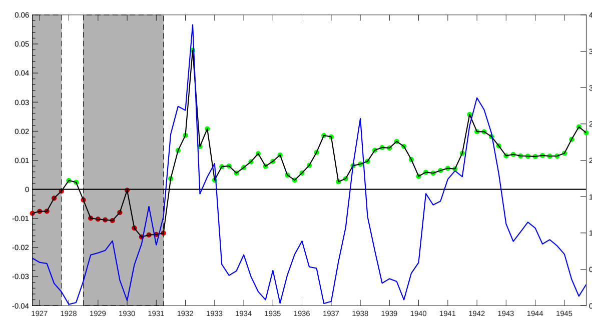
(b) t^{OLS} (Rejection Rate 74%)



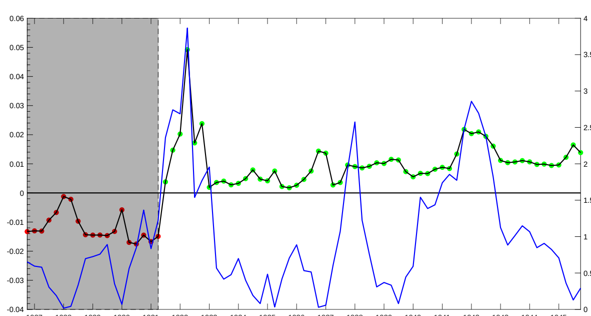
(c) U^{hyb} (Rejection Rate 74%)



(d) W_1^{hyb} (Rejection Rate 78%)



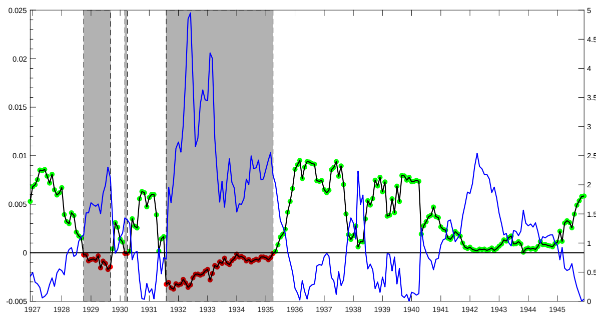
(e) W_2^{hyb} (Rejection Rate 75%)



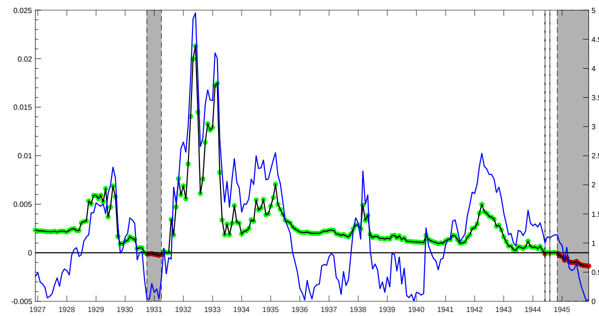
Lower Bound of CI: ——— (Left Axis) ,$|\hat{\theta}|$: ——— (Right Axis)

Figure 8: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Monthly CRSP 1926-1994 (Predictor = $e - p$)

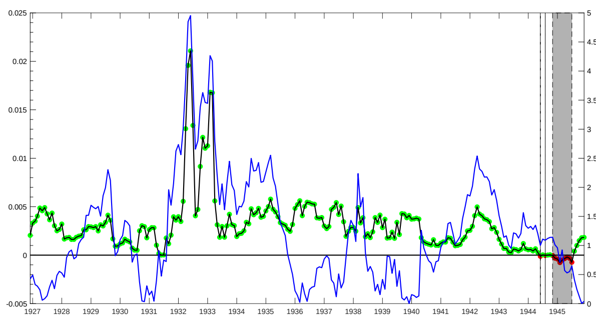
(a) Q^{GLS} (Rejection Rate 74%)



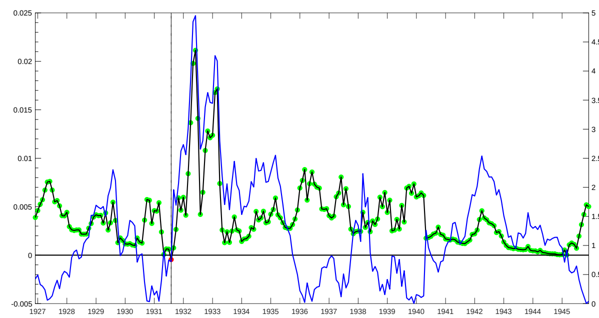
(b) t^{OLS} (Rejection Rate 90%)



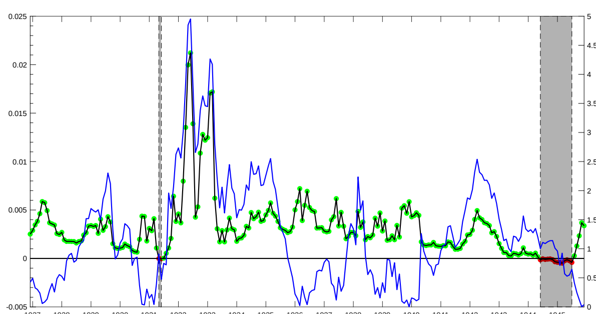
(c) U^{hyb} (Rejection Rate 95%)



(d) W_1^{hyb} (Rejection Rate 100%)



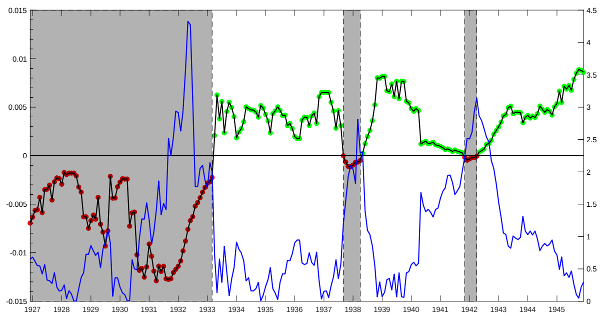
(e) W_2^{hyb} (Rejection Rate 93%)



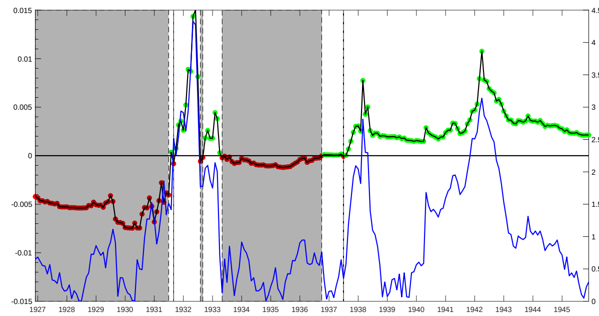
Lower Bound of CI: —— (Left Axis) ,$|\hat{\theta}|$: —— (Right Axis)

Figure 9: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Monthly CRSP 1926-1994 (Predictor = $d - p$)

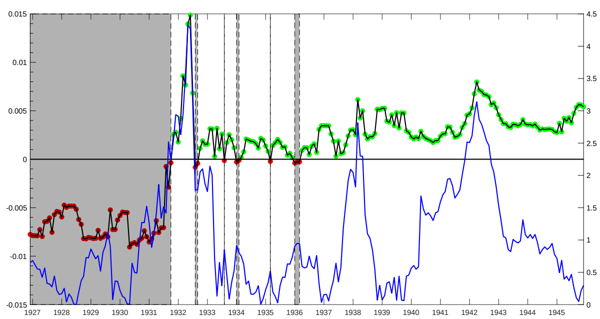
(a) Q^{GLS} (Rejection Rate 61%)



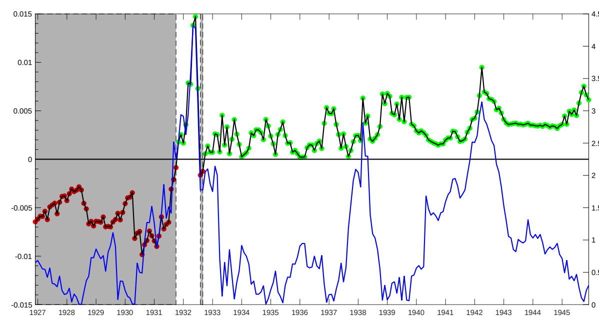
(b) t^{OLS} (Rejection Rate 56%)



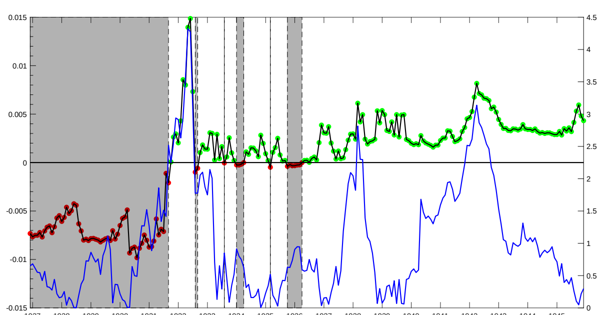
(c) U^{hyb} (Rejection Rate 70%)



(d) W_1^{hyb} (Rejection Rate 73%)



(e) W_2^{hyb} (Rejection Rate 68%)



Lower Bound of CI: ——— (Left Axis) ,$|\hat{\theta}|$: ——— (Right Axis)

Supplementary Material - 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]

[a] Essex Business School, University of Essex

[b] Granger Centre for Time Series Econometrics and School of Economics, University of Nottingham

February 14, 2022

Abstract

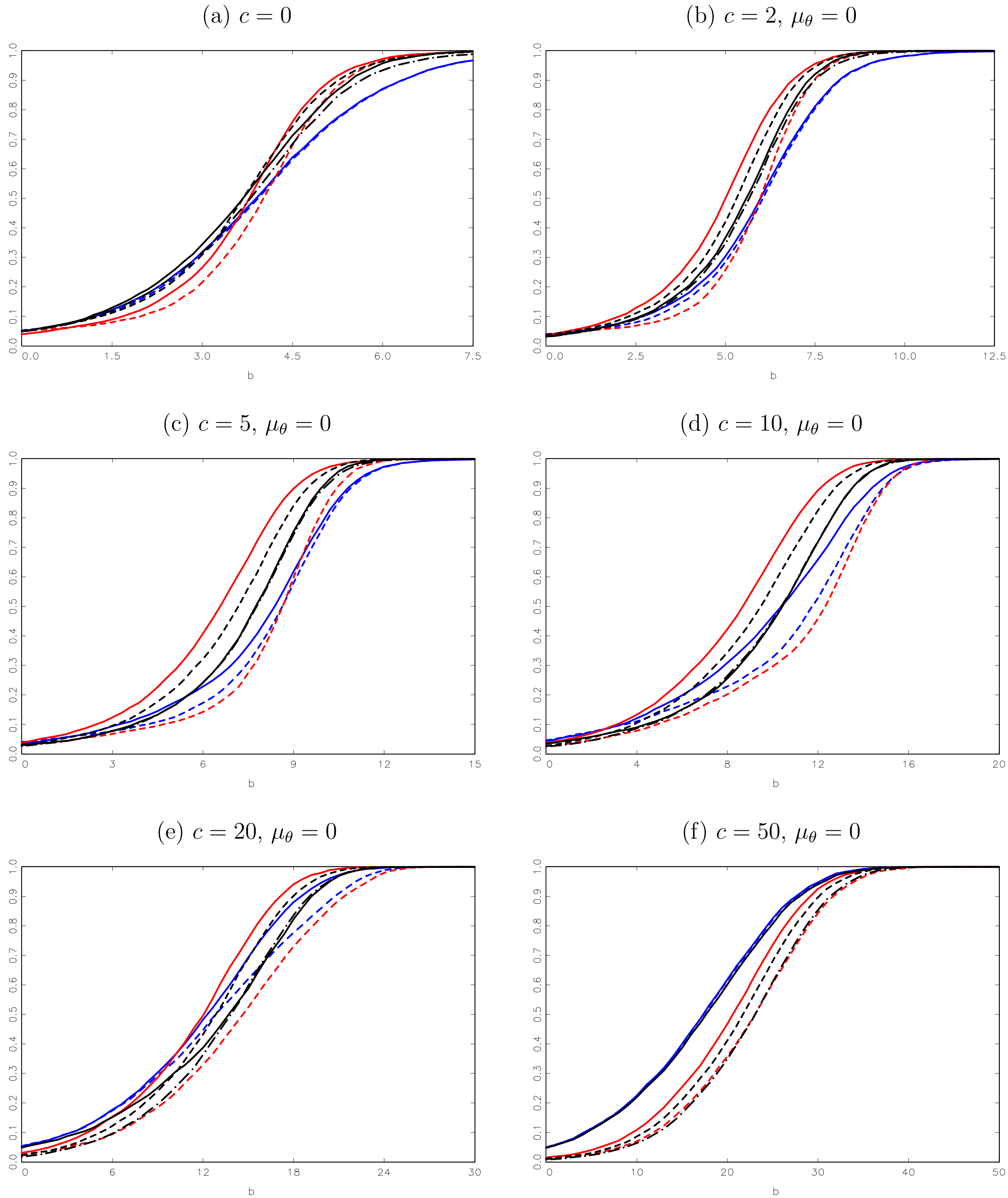
The outline of this supplementary paper is as follows. Section S.1 compares the asymptotic local power of our proposed hybrid test procedures to the extant tests of Campbell and Yogo (2006) [CY] across additional scenarios to those considered in the main paper. Section S.2 reports results from a thorough Monte Carlo simulation exercise examining the finite sample performance of the tests. Section S.3 reports the results of the empirical application of the tests to the full dataset of CY. Finally, Section S.4 provides proofs of the Theorems in the paper.

*We are grateful to Motohiro Yogo for making his Gauss programs to implement the Bonferroni Q and t tests publicly available on his website. Taylor gratefully acknowledges financial support provided by the Economic and Social Research Council of the United Kingdom under research grant ES/R00496X/1.
Address correspondence to: Robert Taylor, Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK. Email: robert.taylor@essex.ac.uk.

S.1 Additional Asymptotic Size and Local Power Comparisons

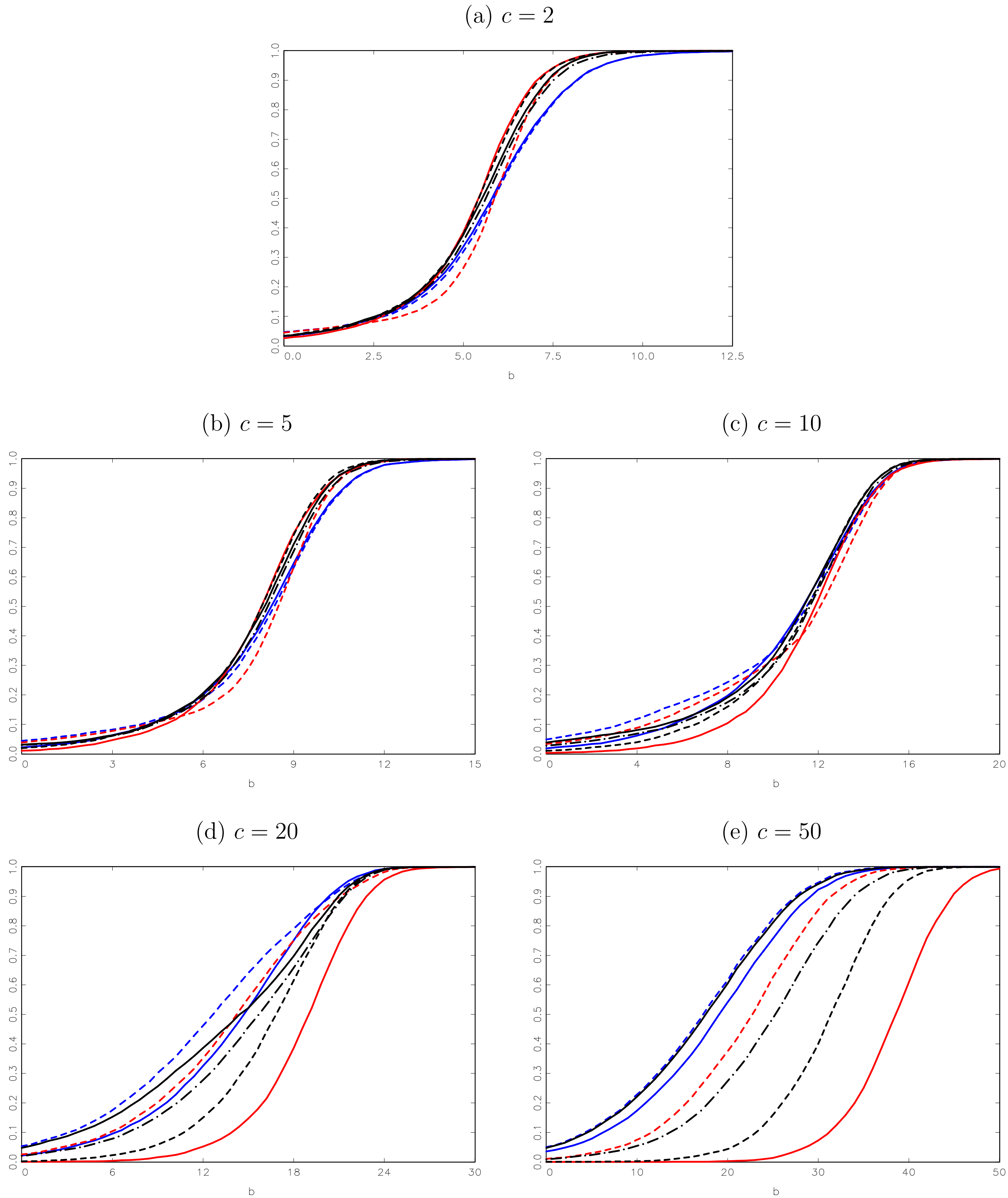
In this section we report additional asymptotic simulation results to those reported in the main paper. We present results for right tailed tests for $c = 0$, and $c = \{2, 5, 10, 20, 50\}$ with $\sigma_\theta^2 = 0$ and $\mu_\theta = \{0, 1, 3\}$, again setting $\sigma_u = \omega_v = 1$. Figures S.1, S.2 and S.3 report results for $\delta = -0.95$, while Figures S.4, S.5 and S.6 report results for $\delta = -0.75$.

Figure S.1: Local Asymptotic Power of Right Tailed Tests - $\delta = -0.95$



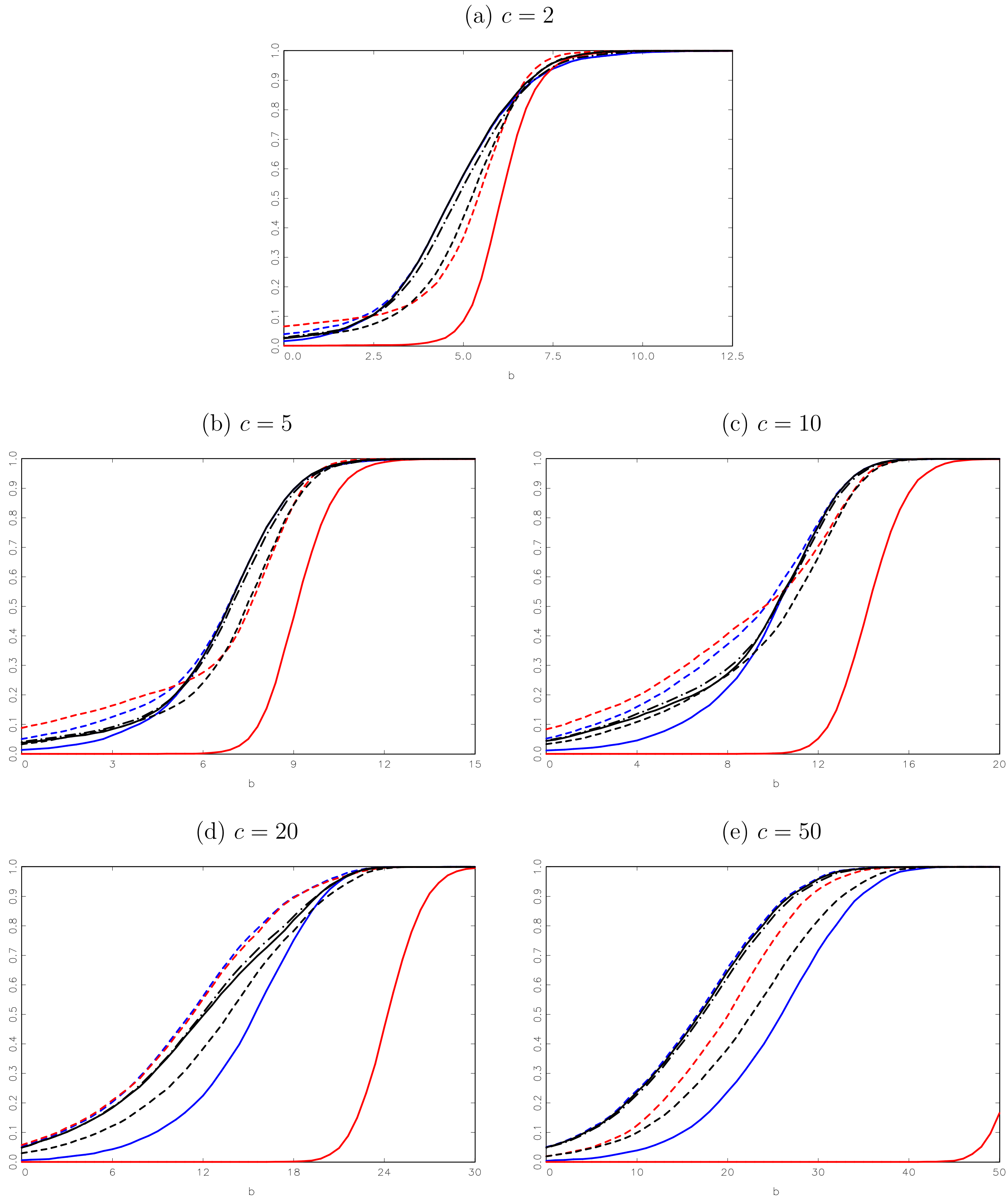
t^{GLS}: ——, t^{OLS}: – –, Q^{GLS}: ——, Q^{OLS}: – –, U^{hyb}——, W_1^{hyb}: – –, W_2^{hyb}: –·–

Figure S.2: Local Asymptotic Power of Right Tailed Tests - $\delta = -0.95$, $\mu_\theta = 1$



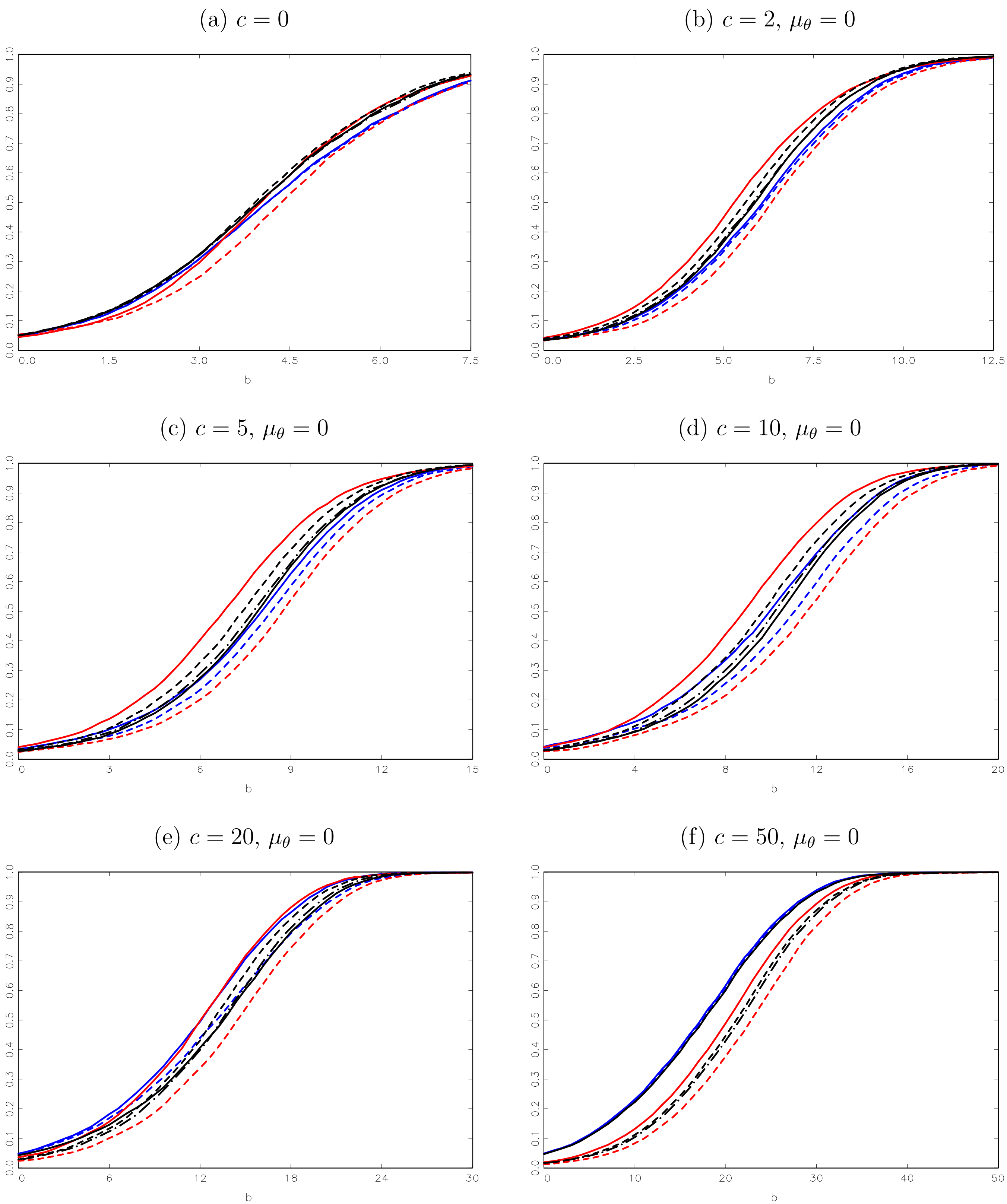
t^{GLS}: ——, t^{OLS}: – – –, Q^{GLS}: ——, Q^{OLS}: – – –, U^{hyb}——, W_1^{hyb}: – –, W_2^{hyb}: –·–

Figure S.3: Local Asymptotic Power of Right Tailed Tests - $\delta = -0.95$, $\mu_\theta = 3$



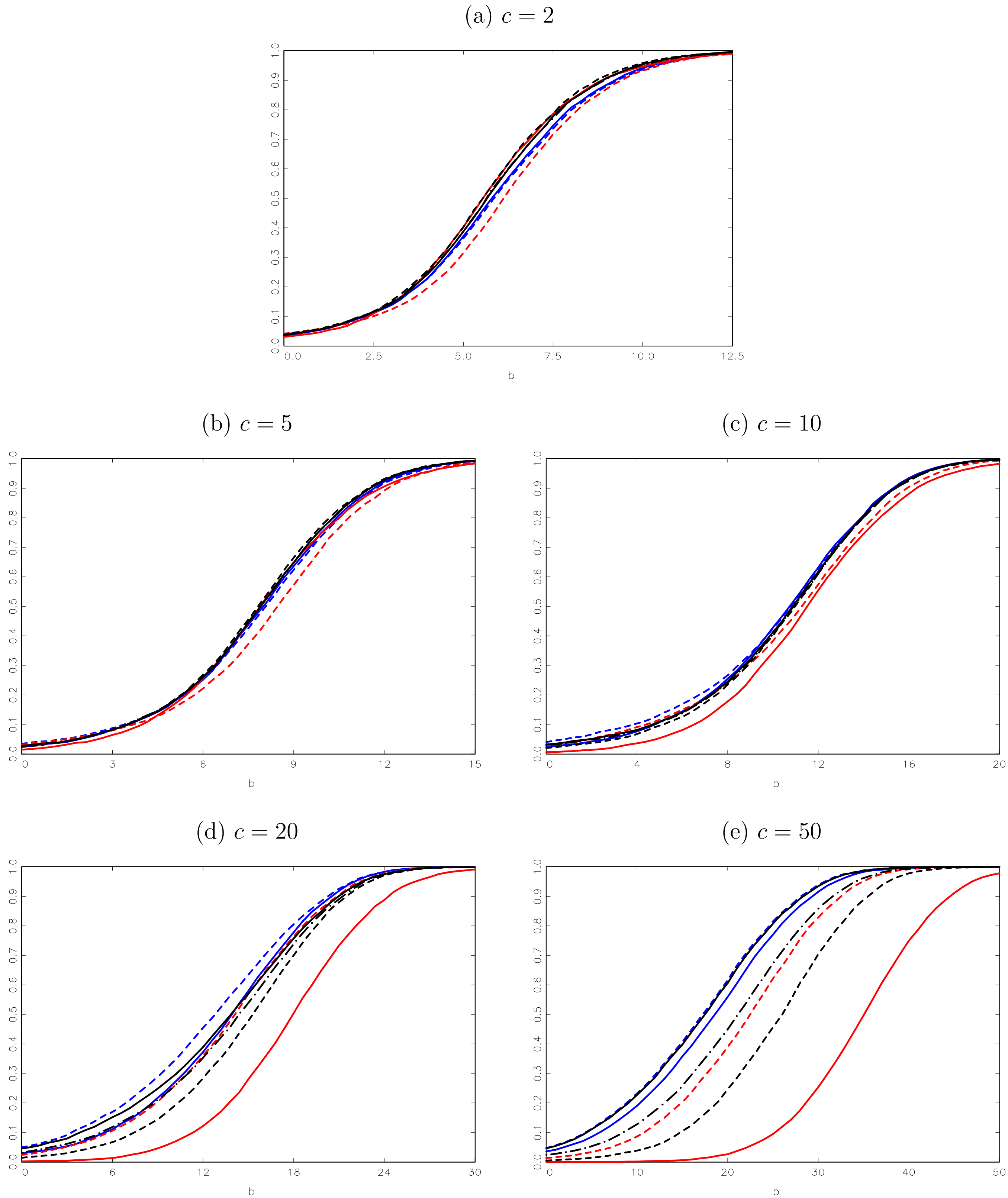
t^{GLS}: ——, t^{OLS}: – – –, Q^{GLS}: ——, Q^{OLS}: – – –, U^{hyb}——, W_1^{hyb}: – –, W_2^{hyb}: –·–

Figure S.4: Local Asymptotic Power of Right Tailed Tests - $\delta = -0.75$



t^{GLS}: ——— , t^{OLS}: – – , Q^{GLS}: ——— , Q^{OLS}: – – , U^{hyb}——— , W_1^{hyb}: – – , W_2^{hyb}: –·–

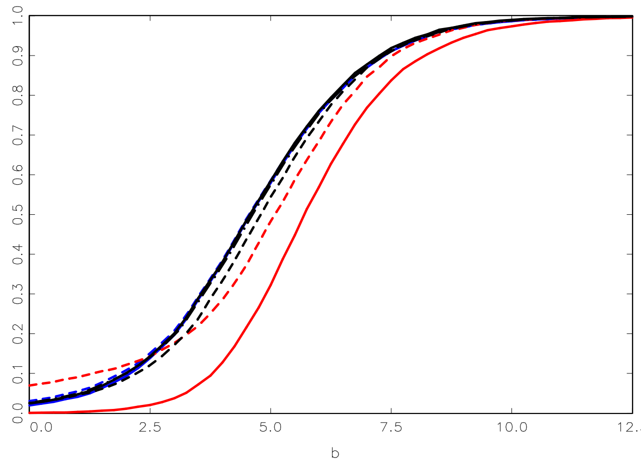
Figure S.5: Local Asymptotic Power of Right Tailed Tests - $\delta = -0.75$, $\mu_\theta = 1$



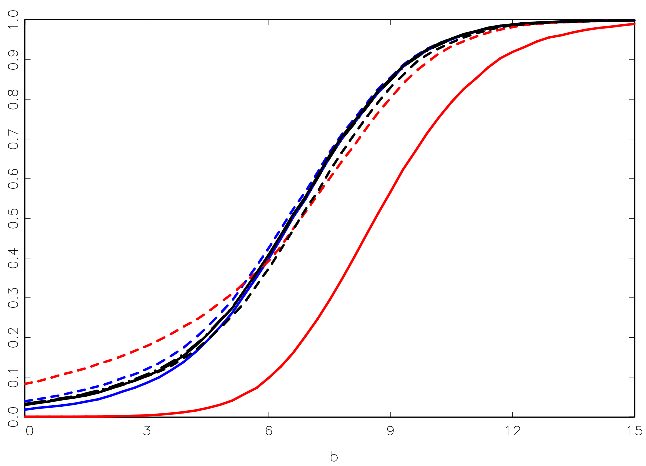
t^{GLS}: ——, t^{OLS}: – –, Q^{GLS}: ——, Q^{OLS}: – –, U^{hyb}——, W_1^{hyb}: – –, W_2^{hyb}: –·–

Figure S.6: Local Asymptotic Power of Right Tailed Tests - $\delta = -0.75$, $\mu_\theta = 3$

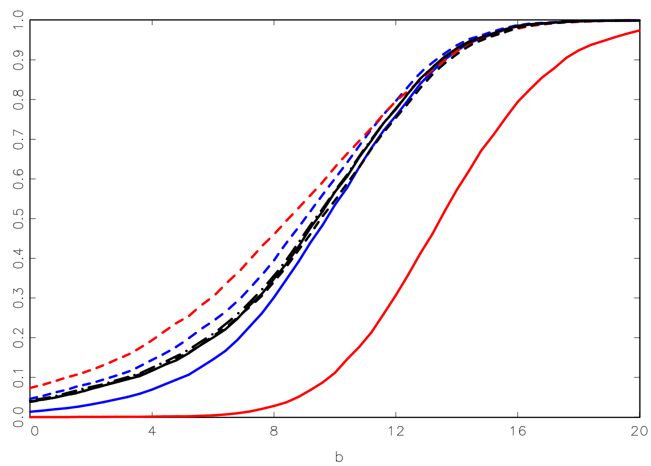
(a) $c = 2$



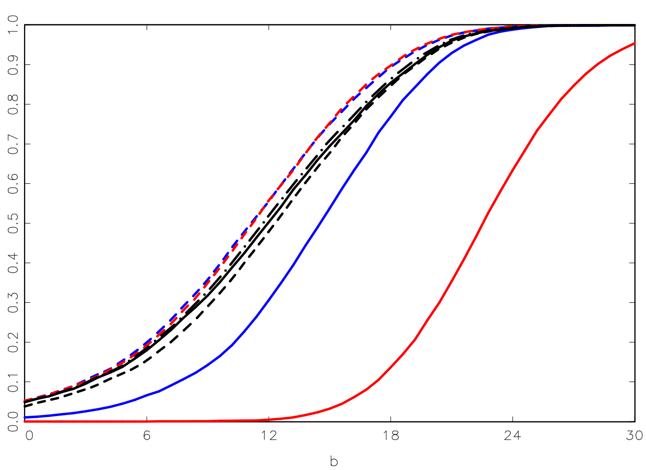
(b) $c = 5$



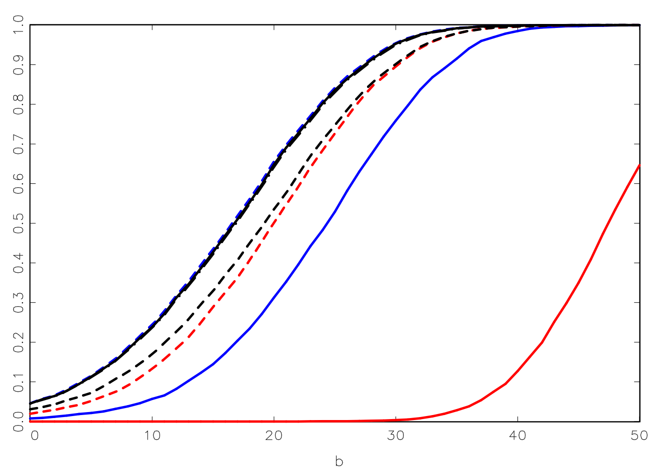
(c) $c = 10$



(d) $c = 20$



(e) $c = 50$



t^{GLS}: ———, t^{OLS}: – – –, Q^{GLS}: ———, Q^{OLS}: – – –, U^{hyb}———, W_1^{hyb}: – –, W_2^{hyb}: –·–

S.2 Finite Sample Performance

In this section we report the finite sample performance of the proposed U^{hyb} and W_{γ}^{hyb}, $\gamma = 1, 2$, procedures relative to the Q^{GLS}, Q^{OLS}, t^{GLS} and t^{OLS} tests. To do so data were generated according to

$$\begin{aligned} r_t &= \alpha + \beta x_{t-1} + u_t, & t &= 1, ..., T \\ x_t &= \mu + w_t, & t &= 0, ..., T \\ w_t &= \rho w_{t-1} + v_t, & t &= 1, ..., T \end{aligned}$$

with $\alpha = \mu = 0$ (without loss of generality), $\rho = 1 - c/T$ and a sample size of $T = 250$. The innovations $\{u_t\}$, $\{v_t\}$ have correlation parameter δ and are drawn from a bivariate normal distribution. We considered $c = 0$ (i.e. Assumption S.1) and a range of $c > 0$ values, $c \in \{2, 5, 10, 20, 50, 100, 250\}$, with the values of $c \leq 50$ chosen to demonstrate the behaviour of the tests when the predictor is a strongly persistent process, and the values of $c \geq 100$ used to illustrate the behaviour of the tests when the predictor is weakly persistent given that $c \geq 100$ implies that $\rho \leq 0.6$. When $c > 0$, the initial condition w_0 is either generated as a $N(0, 1)$, such that Assumption S.2 holds, or as fixed according to Assumption S.3, with $\mu_\theta \in \{1, 3\}$. The DF-OLS and DF-GLS unit root test statistics used to construct the tests for predictability were estimated using a lag length chosen by the Bayes Information Criterion with $p_{\max} = 5$, with this lag length selection method also used for the Dickey-Fuller normalised bias coefficient unit root test statistic. Here and throughout this Monte Carlo simulation exercise, results are reported for right and left tailed tests performed at the nominal 5% level for negative values of δ, noting that results for positive values of δ for right (left) tailed tests are numerically identical to those for negative values of δ for left (right) tailed tests, cf. Remark 3.1. All simulations were performed in Gauss 8.0 using 5,000 Monte Carlo replications.

S.2.1 Finite Sample Size

We begin by setting $\beta = 0$ and examine the finite sample size of right and left tailed tests for predictability for a grid of values of $\delta \in \{-0.95, -0.75, -0.50, -0.25\}$. Table S.1 reports the size of the tests when Assumption S.1 or Assumption S.2 holds. Note that a comparison of the size of the procedures for any given value of c is not appropriate since, while the procedures are constructed to have a maximum asymptotic size of 5% across a range of values of c, the value of c for which the maximum size occurs varies across the different procedures. We see for larger values of c, however, that Q^{GLS} and Q^{OLS} suffer from severe size distortions, with Q^{GLS} exhibiting severe oversize, and Q^{OLS} displaying either severe undersize of oversize depending on the tail under test. This is due to the fact that for larger values of c the predictor will be behaving more like a weakly persistent process in finite samples. We note that, while both U^{hyb} and W_{γ}^{hyb} are a function of the Q^{GLS} test, they do not exhibit any oversize for $c = 250$ which is due to these tests predominantly switching into the standard t-test for larger values of c, although this does lead to some modest oversize for the hybrid tests for $c = 100$ as the standard t test is slightly oversized in this instance.

We next report results when Assumption S.3 holds with $\mu_\theta = 1$, so that the predictor contains a moderately large initial condition, with results presented in Table S.2. We see that U^{hyb}, W_{γ}^{hyb}, t^{OLS} and t^{GLS} are still well size controlled across all values of c and δ. For values of $c \leq 50$, such that the predictor behaves like a local-to-unity process in finite samples, Q^{GLS} can be severely undersized for right tail tests and severely oversized for left tail tests, as predicted by our analysis of the local asymptotic power of this test. We note that U^{hyb} and W_{γ}^{hyb} do not suffer from the severe undersize (oversize) of Q^{GLS} for right (left) tailed testing, despite being a function of Q^{GLS}. This is due to sufficient weight being placed on the size controlled t^{OLS} test in the construction of both U^{hyb} and W_{γ}^{hyb} for right tail testing, and U^{hyb} and W_{γ}^{hyb} reducing to t^{OLS} when testing in the left tail. Both Q^{GLS} and Q^{OLS} continue to suffer severe size distortions for larger values of c where the predictor behaves more like a weakly persistent process in finite samples.

Finally, Table S.3 reports results when Assumption S.3 holds with $\mu_\theta = 3$. Once again the U^{hyb}, W_{γ}^{hyb} and t^{OLS} procedures are very well size-controlled across all values of c. The size issues exhibited by Q^{GLS} when $\mu_\theta = 1$ are further exacerbated when $\mu_\theta = 3$, with the size of this test equal to zero in many instances for right tail tests and equal to almost unity in a number of scenarios for left tail tests. While both t^{GLS} and Q^{OLS} did not suffer from particularly severe size distortions for the lower values of c considered when $\mu_\theta = 1$, we now observe fairly severe size distortions, with t^{GLS} (Q^{OLS}) being substantially undersized (oversized) for right tail tests and oversized (undersized) for left tail tests.

Overall, these simulations demonstrate that the only procedures able to control size for right and left tail testing, across the entire range of c and μ_θ considered, are U^{hyb}, W_γ^{hyb} and t^{OLS}. The Q^{GLS} and Q^{OLS} tests suffer from severe size distortions when the data appears weakly persistent, and are also prone to either severe undersize or oversize when $\mu_\theta \neq 0$ and the predictor is near-integrated. As for t^{GLS}, while reasonably well size controlled for lower values of μ_θ, the test still suffers from non-trivial size distortions for larger values of μ_θ.

S.2.2 Finite Sample Power

We now proceed to examine finite sample power using a grid of 50 values of β for each value of c. We report results for right and left tailed tests with $\delta = -0.95, -0.75$, with these values of δ sufficient to demonstrate the relative power performance of the tests. Results for $\delta = -0.50, -0.25$, available from the authors on request, were also computed but were found to show an identical power ranking between the tests, albeit with the absolute difference in power of the tests reduced to some degree.

Figure S.7 reports the powers of right tailed tests when $\delta = -0.95$ and Assumption S.1 or Assumption S.2 holds. For lower values of $c \leq 20$ we see a similar pattern to what was observed when examining the local asymptotic power of the tests, with Q^{GLS} delivering the best overall power performance, followed by W_1^{hyb}, with U^{hyb} displaying marginally lower power than W_1^{hyb} and W_2^{hyb} having power slightly below that of U^{hyb}. The power of the t^{OLS} test is below that of both the W_γ^{hyb} and U^{hyb} tests, with the power differential larger the lower is the value of c. For $c = 50$, we observe U^{hyb} outperforming W_1^{hyb} and W_2^{hyb}, while Q^{GLS} suffers from poor power. The power of the new hybrid tests coincide and dominate that of all other tests when $c = 100, 250$ due to these tests switching into the standard t-test in a majority of replications, which is optimal for weakly persistent predictors. Figures S.8 and S.9 report powers of right tailed tests when $\delta = -0.95$ and Assumption S.3 holds with $\mu_\theta = 1$ and $\mu_\theta = 3$, respectively. In these larger initial condition cases, we observe the same pattern as in the corresponding limit results, with the $\mu_\theta = 0$ power advantages of Q^{GLS} almost entirely eliminated and the hybrid procedures displaying excellent power due to their close correspondence to the power profile of t^{OLS} for lower values of c and the fact that they predominantly switch into the optimal standard t-test for larger values of c.

Figures S.10, S.11 and S.12 report results for right tailed tests when $\delta = -0.75$ and $\mu_\theta = 0$, 1 and 3, respectively, and similar comments apply to those for Figures S.7, S.8 and S.9.

Figure S.13 reports powers of left tailed tests when $\delta = -0.95$ and Assumption S.1 or Assumption S.2 holds, and Figures S.14 and S.15 report powers of left tailed tests when $\delta = -0.95$ and Assumption S.3 holds with $\mu_\theta = 1$ and $\mu_\theta = 3$, respectively. First we note that the power functions for the W_γ^{hyb} and U^{hyb} procedures are identical, given that in this left tailed testing context, the procedures reduce to either t^{OLS} or the standard t-test. When $\mu_\theta = 0$, Q^{GLS} delivers the best overall power performance only for the relatively small values of $c = 0, 2$, mirroring the results of our local asymptotic power analysis; for larger c, the best performing procedures are W_γ^{hyb}, U^{hyb} and t^{OLS}. The small differences between t^{OLS} and the W_γ^{hyb} and U^{hyb} procedures arise from occasions where the hybrid procedures switch into the standard t-test for larger c. While it may appear strange that the t^{OLS} test displays slightly higher power than our hybrid test procedures for $c = 250$ when the hybrid procedures will be running off of the optimal standard t test more often than not, we note that the t^{OLS} test displays a modest degree of oversize in this particular scenario. When $\mu_\theta = 1$ or $\mu_\theta = 3$, substantial oversize emerges in Q^{GLS}, as was observed in the local power results. The hybrid procedures W_γ^{hyb} and U^{hyb} coincide with t^{OLS} in most cases, with attractive power results displayed across the full range of c. Figures S.16, S.17 and S.18 present the results corresponding to Figures S.13, S.14 and S.15, respectively, for the case $\delta = -0.75$, and similar comments apply.

Overall the results in this Monte Carlo exercise suggest that, in common with the local asymptotic power results, the best overall power performance is displayed by the new hybrid procedure U^{hyb}, with W_1^{hyb} and W_2^{hyb} also performing very well.

Table S.1: Finite Sample Size, $T = 250$, $\mu_\theta = 0$.

(a) Right Tailed Tests									(b) Left Tailed Tests								
c	δ	t^{GLS}	t^{OLS}	Q^{GLS}	Q^{OLS}	U^{hyb}	W_1^{hyb}	W_2^{hyb}	c	δ	t^{GLS}	t^{OLS}	Q^{GLS}	Q^{OLS}	U^{hyb}	W_1^{hyb}	W_2^{hyb}
0	-0.95	0.049	0.052	0.044	0.052	0.048	0.054	0.046	0	-0.95	0.000	0.003	0.009	0.017	0.003	0.003	0.003
	-0.75	0.050	0.052	0.048	0.054	0.051	0.055	0.052		-0.75	0.005	0.007	0.009	0.009	0.007	0.007	0.007
	-0.50	0.048	0.051	0.052	0.054	0.048	0.052	0.050		-0.50	0.023	0.020	0.017	0.012	0.020	0.020	0.020
	-0.25	0.050	0.051	0.054	0.054	0.051	0.048	0.049		-0.25	0.039	0.028	0.030	0.021	0.028	0.028	0.028
2	-0.95	0.039	0.042	0.048	0.041	0.036	0.043	0.036	2	-0.95	0.004	0.011	0.008	0.027	0.011	0.011	0.011
	-0.75	0.038	0.037	0.046	0.038	0.034	0.041	0.035		-0.75	0.014	0.018	0.007	0.014	0.018	0.018	0.018
	-0.50	0.040	0.039	0.050	0.039	0.036	0.040	0.037		-0.50	0.032	0.028	0.015	0.014	0.028	0.028	0.028
	-0.25	0.040	0.042	0.052	0.043	0.040	0.041	0.040		-0.25	0.044	0.034	0.027	0.024	0.034	0.034	0.034
5	-0.95	0.040	0.046	0.045	0.034	0.034	0.041	0.035	5	-0.95	0.017	0.033	0.012	0.026	0.033	0.033	0.033
	-0.75	0.040	0.034	0.045	0.029	0.029	0.037	0.033		-0.75	0.031	0.034	0.008	0.015	0.034	0.034	0.034
	-0.50	0.037	0.034	0.050	0.032	0.030	0.037	0.031		-0.50	0.040	0.037	0.015	0.015	0.037	0.037	0.037
	-0.25	0.039	0.041	0.053	0.041	0.039	0.041	0.038		-0.25	0.046	0.041	0.026	0.026	0.041	0.041	0.041
10	-0.95	0.042	0.047	0.044	0.035	0.039	0.039	0.038	10	-0.95	0.037	0.045	0.018	0.023	0.045	0.045	0.045
	-0.75	0.043	0.041	0.043	0.030	0.033	0.039	0.034		-0.75	0.040	0.042	0.011	0.014	0.042	0.042	0.042
	-0.50	0.039	0.036	0.045	0.031	0.031	0.034	0.032		-0.50	0.050	0.048	0.016	0.016	0.048	0.048	0.048
	-0.25	0.037	0.038	0.052	0.040	0.035	0.038	0.035		-0.25	0.051	0.047	0.028	0.027	0.047	0.047	0.047
20	-0.95	0.046	0.051	0.036	0.037	0.048	0.035	0.036	20	-0.95	0.045	0.049	0.036	0.022	0.049	0.049	0.049
	-0.75	0.044	0.045	0.034	0.031	0.041	0.034	0.036		-0.75	0.047	0.048	0.020	0.014	0.048	0.048	0.048
	-0.50	0.044	0.044	0.038	0.032	0.039	0.034	0.033		-0.50	0.047	0.047	0.022	0.015	0.047	0.047	0.047
	-0.25	0.039	0.043	0.045	0.039	0.038	0.035	0.036		-0.25	0.051	0.049	0.029	0.026	0.049	0.049	0.049
50	-0.95	0.044	0.048	0.032	0.048	0.056	0.052	0.050	50	-0.95	0.048	0.048	0.152	0.012	0.048	0.048	0.048
	-0.75	0.041	0.044	0.028	0.035	0.048	0.038	0.040		-0.75	0.048	0.046	0.088	0.009	0.046	0.046	0.046
	-0.50	0.044	0.045	0.030	0.036	0.046	0.034	0.035		-0.50	0.052	0.050	0.056	0.011	0.050	0.050	0.050
	-0.25	0.044	0.046	0.035	0.044	0.046	0.036	0.039		-0.25	0.052	0.050	0.045	0.021	0.050	0.050	0.050
100	-0.95	0.038	0.041	0.048	0.115	0.060	0.060	0.060	100	-0.95	0.051	0.049	0.322	0.004	0.033	0.033	0.033
	-0.75	0.040	0.043	0.040	0.082	0.057	0.057	0.057		-0.75	0.052	0.051	0.241	0.002	0.037	0.037	0.037
	-0.50	0.041	0.045	0.032	0.063	0.053	0.053	0.053		-0.50	0.054	0.053	0.154	0.006	0.042	0.042	0.042
	-0.25	0.044	0.046	0.032	0.056	0.052	0.052	0.052		-0.25	0.052	0.052	0.081	0.017	0.047	0.047	0.047
250	-0.95	0.035	0.036	0.275	0.899	0.053	0.053	0.053	250	-0.95	0.070	0.071	0.454	0.014	0.046	0.046	0.046
	-0.75	0.035	0.035	0.269	0.770	0.049	0.049	0.049		-0.75	0.066	0.066	0.414	0.010	0.046	0.046	0.046
	-0.50	0.037	0.038	0.203	0.604	0.047	0.047	0.047		-0.50	0.061	0.061	0.350	0.006	0.050	0.050	0.050
	-0.25	0.039	0.039	0.092	0.257	0.044	0.044	0.044		-0.25	0.053	0.053	0.212	0.005	0.048	0.048	0.048

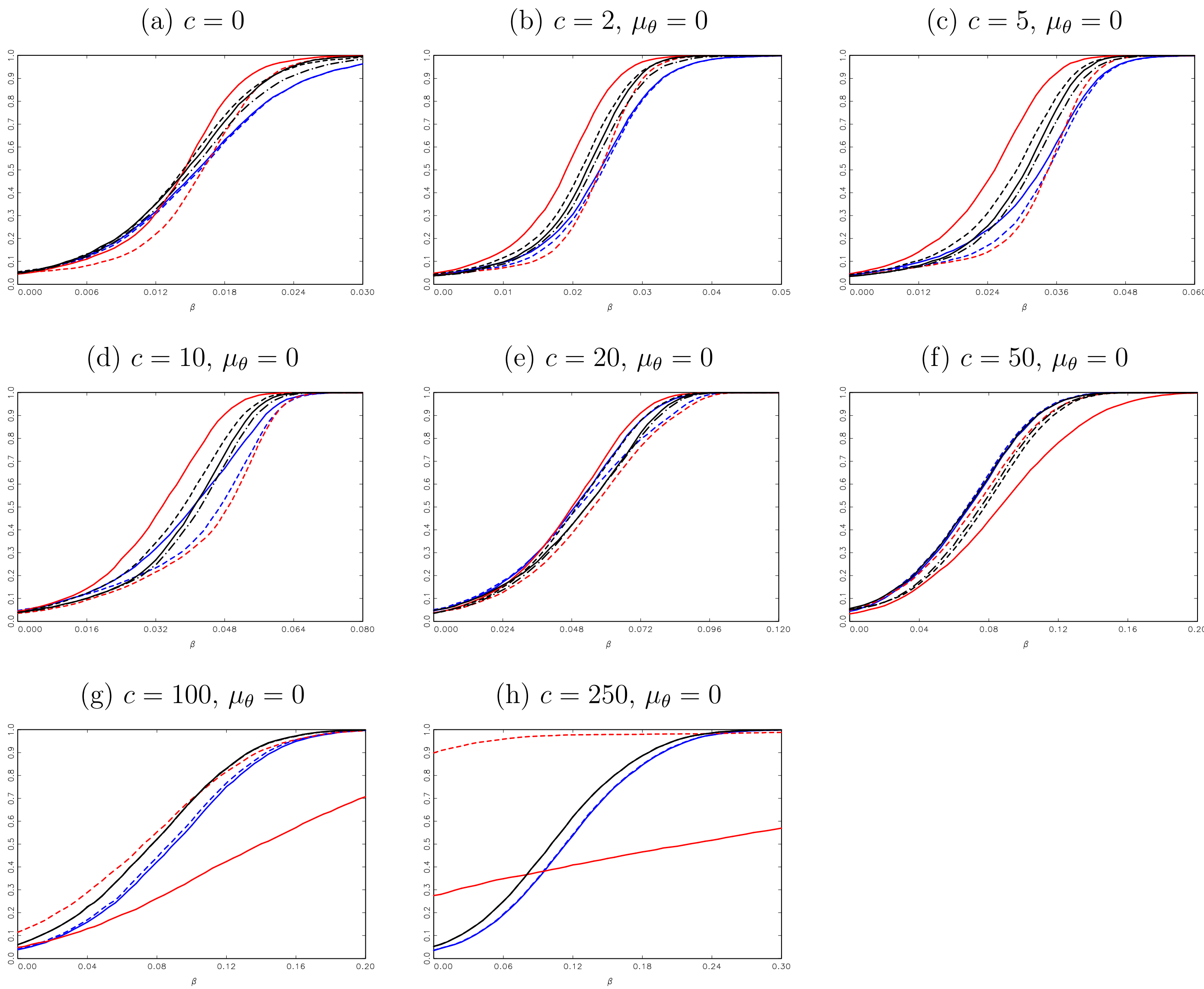
Table S.2: Finite Sample Size, $T = 250$, $\mu_\theta = 1$.

		(a) Right Tailed Tests									(b) Left Tailed Tests						
c	δ	t^{GLS}	t^{OLS}	Q^{GLS}	Q^{OLS}	U^{hyb}	W_1^{hyb}	W_2^{hyb}	c	δ	t^{GLS}	t^{OLS}	Q^{GLS}	Q^{OLS}	U^{hyb}	W_1^{hyb}	W_2^{hyb}
2	-0.95	0.027	0.037	0.028	0.039	0.027	0.034	0.027	2	-0.95	0.005	0.013	0.032	0.034	0.013	0.013	0.013
	-0.75	0.030	0.032	0.032	0.034	0.028	0.033	0.029		-0.75	0.015	0.019	0.019	0.015	0.019	0.019	0.019
	-0.50	0.035	0.036	0.044	0.037	0.033	0.037	0.035		-0.50	0.036	0.031	0.022	0.014	0.031	0.031	0.031
	-0.25	0.041	0.042	0.049	0.044	0.042	0.041	0.042		-0.25	0.050	0.036	0.034	0.023	0.036	0.036	0.036
5	-0.95	0.020	0.039	0.014	0.034	0.026	0.028	0.028	5	-0.95	0.018	0.038	0.064	0.026	0.038	0.038	0.038
	-0.75	0.024	0.029	0.017	0.031	0.024	0.025	0.025		-0.75	0.033	0.037	0.032	0.016	0.037	0.037	0.037
	-0.50	0.029	0.032	0.028	0.033	0.028	0.030	0.029		-0.50	0.049	0.041	0.032	0.016	0.041	0.041	0.041
	-0.25	0.038	0.041	0.042	0.043	0.038	0.037	0.037		-0.25	0.054	0.044	0.041	0.027	0.044	0.044	0.044
10	-0.95	0.019	0.045	0.006	0.036	0.036	0.025	0.034	10	-0.95	0.040	0.048	0.111	0.021	0.048	0.048	0.048
	-0.75	0.025	0.037	0.009	0.033	0.031	0.026	0.031		-0.75	0.046	0.043	0.057	0.013	0.043	0.043	0.043
	-0.50	0.028	0.035	0.018	0.034	0.030	0.028	0.031		-0.50	0.054	0.048	0.046	0.014	0.048	0.048	0.048
	-0.25	0.033	0.039	0.034	0.042	0.037	0.034	0.035		-0.25	0.055	0.049	0.050	0.027	0.049	0.049	0.049
20	-0.95	0.027	0.049	0.003	0.037	0.045	0.019	0.035	20	-0.95	0.050	0.047	0.222	0.019	0.047	0.047	0.047
	-0.75	0.031	0.044	0.004	0.032	0.041	0.022	0.034		-0.75	0.052	0.048	0.103	0.013	0.048	0.048	0.048
	-0.50	0.033	0.042	0.010	0.034	0.037	0.025	0.034		-0.50	0.056	0.049	0.072	0.013	0.049	0.049	0.049
	-0.25	0.031	0.045	0.025	0.041	0.038	0.031	0.035		-0.25	0.053	0.049	0.059	0.026	0.049	0.049	0.049
50	-0.95	0.035	0.048	0.000	0.049	0.053	0.043	0.046	50	-0.95	0.054	0.050	0.525	0.011	0.050	0.050	0.050
	-0.75	0.034	0.044	0.000	0.038	0.048	0.031	0.038		-0.75	0.050	0.046	0.252	0.008	0.046	0.046	0.046
	-0.50	0.038	0.043	0.003	0.037	0.044	0.024	0.036		-0.50	0.054	0.049	0.138	0.011	0.049	0.049	0.049
	-0.25	0.040	0.046	0.013	0.043	0.046	0.030	0.040		-0.25	0.052	0.049	0.087	0.023	0.049	0.049	0.049
100	-0.95	0.037	0.040	0.000	0.116	0.061	0.061	0.061	100	-0.95	0.053	0.051	0.791	0.004	0.033	0.033	0.033
	-0.75	0.036	0.042	0.000	0.085	0.055	0.055	0.055		-0.75	0.052	0.049	0.484	0.003	0.036	0.036	0.036
	-0.50	0.038	0.043	0.001	0.062	0.054	0.054	0.054		-0.50	0.053	0.051	0.247	0.005	0.040	0.040	0.040
	-0.25	0.042	0.046	0.007	0.055	0.051	0.051	0.051		-0.25	0.052	0.051	0.119	0.017	0.045	0.045	0.045
250	-0.95	0.035	0.036	0.000	0.898	0.053	0.053	0.053	250	-0.95	0.071	0.071	0.969	0.015	0.047	0.047	0.047
	-0.75	0.035	0.035	0.000	0.771	0.049	0.049	0.049		-0.75	0.065	0.065	0.855	0.010	0.047	0.047	0.047
	-0.50	0.039	0.039	0.000	0.603	0.047	0.047	0.047		-0.50	0.060	0.060	0.546	0.006	0.049	0.049	0.049
	-0.25	0.039	0.039	0.003	0.255	0.044	0.044	0.044		-0.25	0.054	0.054	0.225	0.005	0.047	0.047	0.047

Table S.3: Finite Sample Size, $T = 250$, $\mu_\theta = 3$.

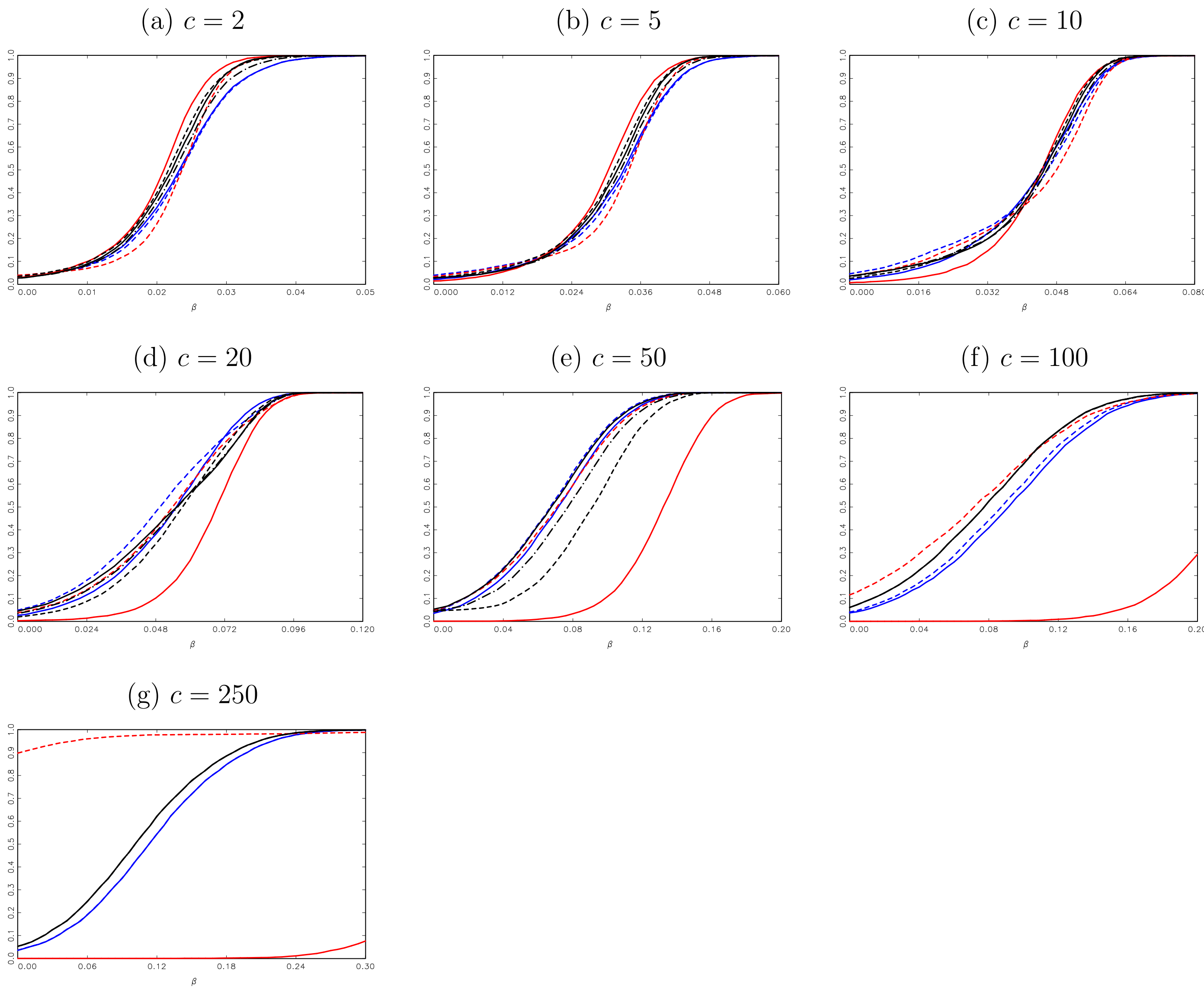
		(a) Right Tailed Tests									(b) Left Tailed Tests						
c	δ	t^{GLS}	t^{OLS}	Q^{GLS}	Q^{OLS}	U^{hyb}	W_1^{hyb}	W_2^{hyb}	c	δ	t^{GLS}	t^{OLS}	Q^{GLS}	Q^{OLS}	U^{hyb}	W_1^{hyb}	W_2^{hyb}
2	-0.95	0.018	0.040	0.001	0.065	0.030	0.034	0.031	2	-0.95	0.021	0.047	0.315	0.061	0.047	0.047	0.047
	-0.75	0.018	0.030	0.001	0.067	0.026	0.026	0.027		-0.75	0.034	0.046	0.141	0.035	0.046	0.046	0.046
	-0.50	0.025	0.029	0.008	0.061	0.027	0.026	0.027		-0.50	0.057	0.049	0.090	0.024	0.049	0.049	0.049
	-0.25	0.035	0.037	0.026	0.052	0.035	0.034	0.036		-0.25	0.062	0.042	0.066	0.024	0.042	0.042	0.042
5	-0.95	0.010	0.046	0.000	0.089	0.035	0.037	0.038	5	-0.95	0.030	0.058	0.677	0.018	0.058	0.058	0.058
	-0.75	0.010	0.033	0.000	0.081	0.025	0.027	0.028		-0.75	0.046	0.051	0.288	0.012	0.051	0.051	0.051
	-0.50	0.018	0.030	0.003	0.067	0.024	0.025	0.025		-0.50	0.072	0.047	0.156	0.008	0.047	0.047	0.047
	-0.25	0.033	0.037	0.018	0.058	0.034	0.033	0.035		-0.25	0.071	0.046	0.094	0.016	0.046	0.046	0.046
10	-0.95	0.004	0.047	0.000	0.094	0.041	0.039	0.043	10	-0.95	0.043	0.051	0.958	0.005	0.051	0.051	0.051
	-0.75	0.007	0.041	0.000	0.080	0.034	0.034	0.035		-0.75	0.056	0.045	0.516	0.004	0.045	0.045	0.045
	-0.50	0.016	0.036	0.001	0.065	0.027	0.030	0.031		-0.50	0.080	0.048	0.239	0.005	0.048	0.048	0.048
	-0.25	0.030	0.037	0.011	0.053	0.034	0.033	0.034		-0.25	0.069	0.049	0.118	0.013	0.049	0.049	0.049
20	-0.95	0.003	0.043	0.000	0.085	0.040	0.036	0.041	20	-0.95	0.060	0.047	1.000	0.004	0.047	0.047	0.047
	-0.75	0.005	0.045	0.000	0.069	0.044	0.037	0.044		-0.75	0.079	0.044	0.814	0.003	0.044	0.044	0.044
	-0.50	0.013	0.044	0.000	0.059	0.039	0.039	0.040		-0.50	0.083	0.048	0.394	0.005	0.048	0.048	0.048
	-0.25	0.027	0.045	0.007	0.054	0.039	0.039	0.041		-0.25	0.067	0.049	0.165	0.017	0.049	0.049	0.049
50	-0.95	0.005	0.044	0.000	0.097	0.048	0.044	0.048	50	-0.95	0.087	0.053	1.000	0.004	0.053	0.053	0.053
	-0.75	0.008	0.045	0.000	0.074	0.048	0.040	0.048		-0.75	0.079	0.047	0.992	0.003	0.047	0.047	0.047
	-0.50	0.012	0.044	0.000	0.060	0.045	0.039	0.045		-0.50	0.075	0.051	0.747	0.006	0.051	0.051	0.051
	-0.25	0.023	0.045	0.002	0.054	0.045	0.040	0.044		-0.25	0.064	0.049	0.293	0.016	0.049	0.049	0.049
100	-0.95	0.012	0.042	0.000	0.184	0.060	0.060	0.060	100	-0.95	0.078	0.053	0.999	0.003	0.034	0.034	0.034
	-0.75	0.014	0.041	0.000	0.127	0.060	0.060	0.060		-0.75	0.073	0.051	0.998	0.002	0.036	0.036	0.036
	-0.50	0.022	0.042	0.000	0.090	0.052	0.052	0.052		-0.50	0.067	0.050	0.956	0.002	0.042	0.042	0.042
	-0.25	0.022	0.046	0.000	0.067	0.051	0.051	0.051		-0.25	0.060	0.050	0.489	0.012	0.045	0.045	0.045
250	-0.95	0.021	0.034	0.000	0.931	0.053	0.053	0.053	250	-0.95	0.081	0.072	1.000	0.014	0.051	0.051	0.051
	-0.75	0.025	0.035	0.000	0.835	0.046	0.046	0.046		-0.75	0.073	0.065	0.998	0.009	0.046	0.046	0.046
	-0.50	0.030	0.039	0.000	0.672	0.047	0.047	0.047		-0.50	0.065	0.059	0.995	0.005	0.048	0.048	0.048
	-0.25	0.035	0.040	0.000	0.297	0.046	0.046	0.046		-0.25	0.056	0.054	0.891	0.005	0.048	0.048	0.048

Figure S.7: Finite Sample Power Right Tailed Tests - $T = 250$, $\delta = -0.95$



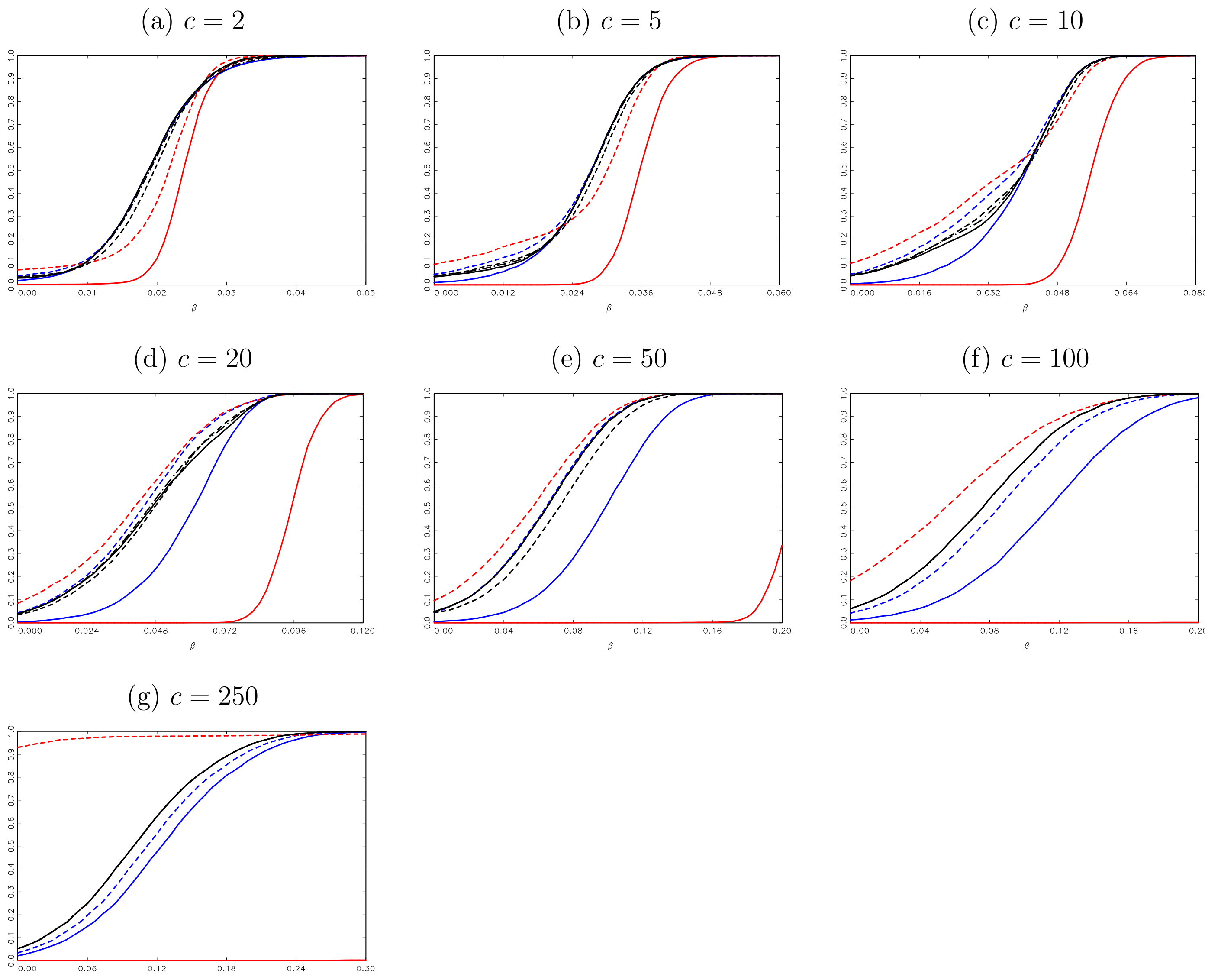
t^{GLS}: ——— , t^{OLS}: – – , Q^{GLS}: ———, Q^{OLS}: – –, U^{hyb}———, W_1^{hyb}: – –, W_2^{hyb}: –·–

Figure S.8: Finite Sample Power Right Tailed Tests - $T = 250$, $\mu_\theta = 1$, $\delta = -0.95$



t^{GLS}: ——, t^{OLS}: – –, Q^{GLS}: ——, Q^{OLS}: – –, U^{hyb}——, W_1^{hyb}: – –, W_2^{hyb}: –·–

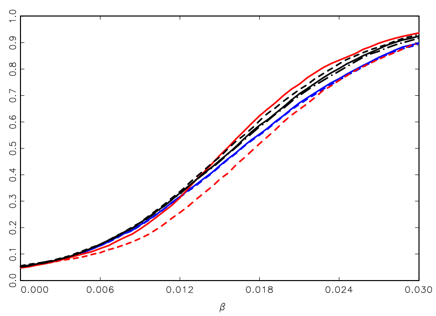
Figure S.9: Finite Sample Power Right Tailed Tests - $T = 250$, $\mu_\theta = 3$, $\delta = -0.95$



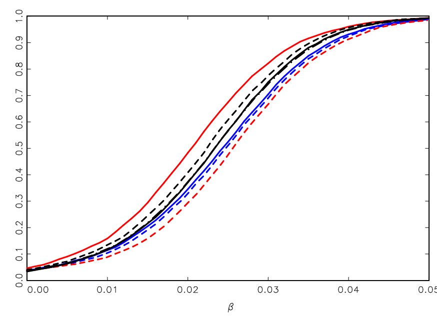
t^{GLS}: ——— , t^{OLS}: – – , Q^{GLS}: ———, Q^{OLS}: – –, U^{hyb}———, W_1^{hyb}: – –, W_2^{hyb}: –•–

Figure S.10: Finite Sample Power Right Tailed Tests - $T = 250$, $\delta = -0.75$

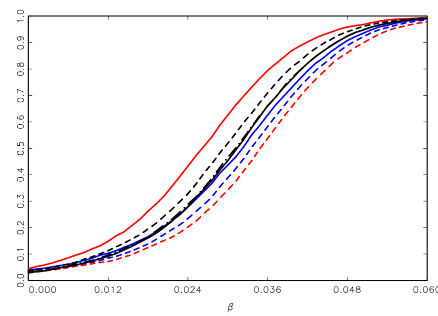
(a) $c = 0$



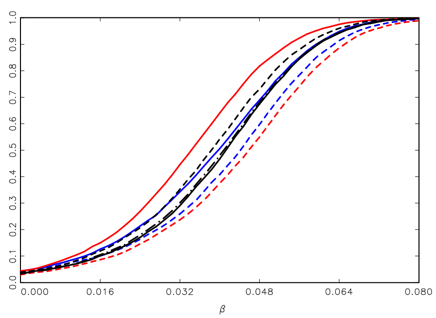
(b) $c = 2$, $\mu_\theta = 0$



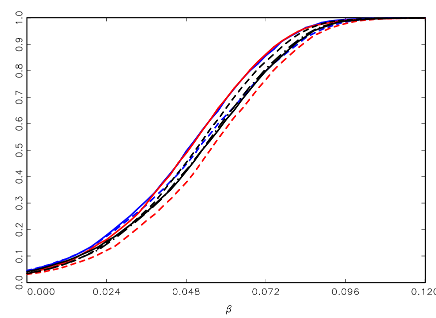
(c) $c = 5$, $\mu_\theta = 0$



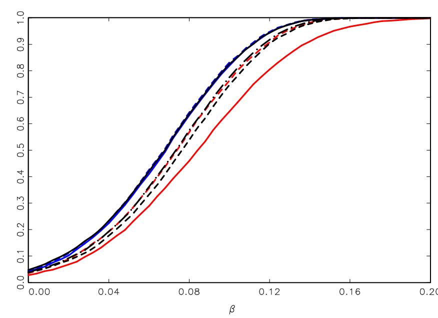
(d) $c = 10$, $\mu_\theta = 0$



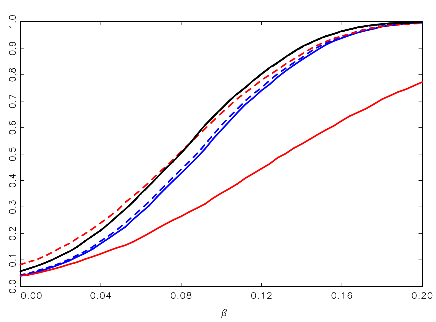
(e) $c = 20$, $\mu_\theta = 0$



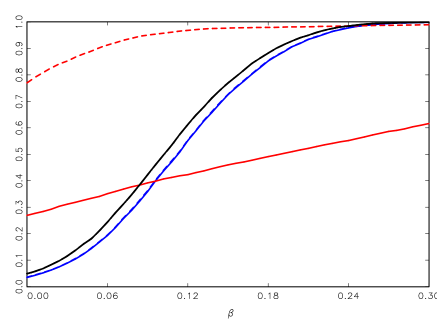
(f) $c = 50$, $\mu_\theta = 0$



(g) $c = 100$, $\mu_\theta = 0$

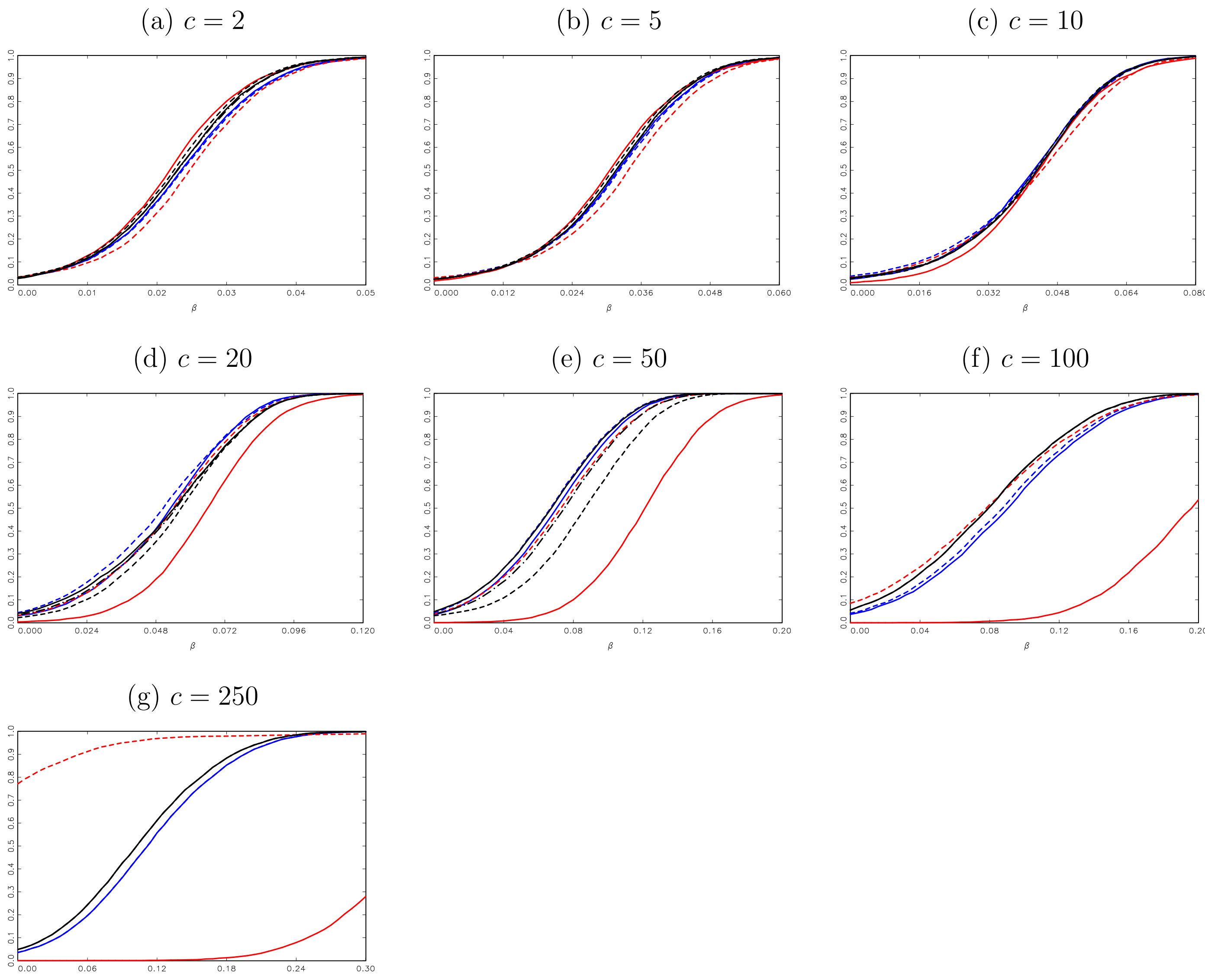


(h) $c = 250$, $\mu_\theta = 0$



t^{GLS}: ——— , t^{OLS}: – – , Q^{GLS}: ———, Q^{OLS}: – –, U^{hyb}———, W_1^{hyb}: – –, W_2^{hyb}: –•–

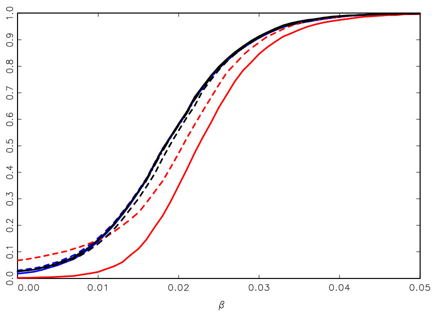
Figure S.11: Finite Sample Power Right Tailed Tests - $T = 250$, $\mu_\theta = 1$, $\delta = -0.75$



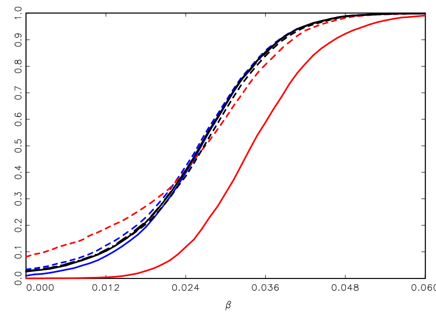
t^{GLS}: ——— , t^{OLS}: – – , Q^{GLS}: ———, Q^{OLS}: – –, U^{hyb}———, W_1^{hyb}: – –, W_2^{hyb}: –·–

Figure S.12: Finite Sample Power Right Tailed Tests - $T = 250$, $\mu_\theta = 3$, $\delta = -0.75$

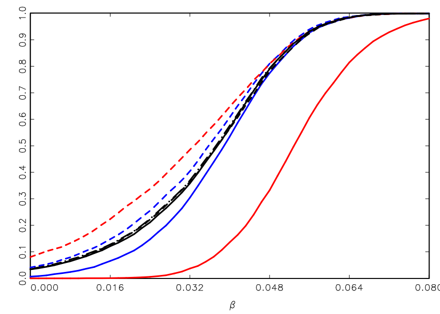
(a) $c = 2$



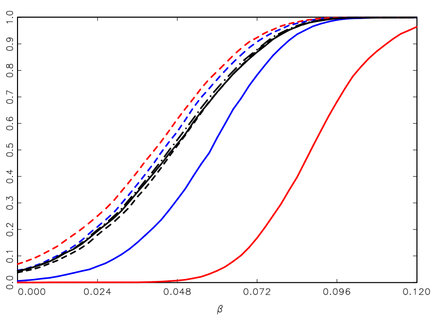
(b) $c = 5$



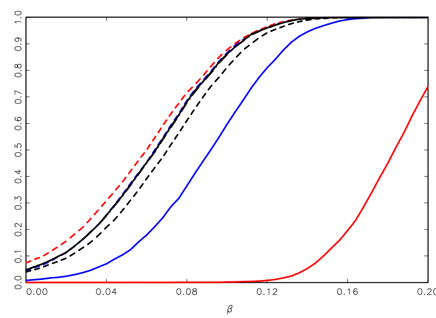
(c) $c = 10$



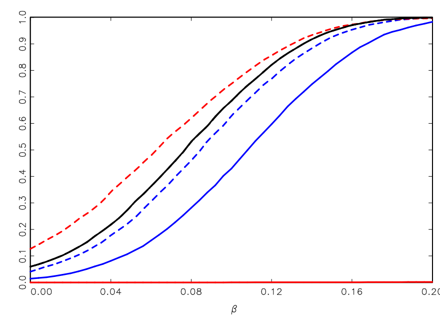
(d) $c = 20$



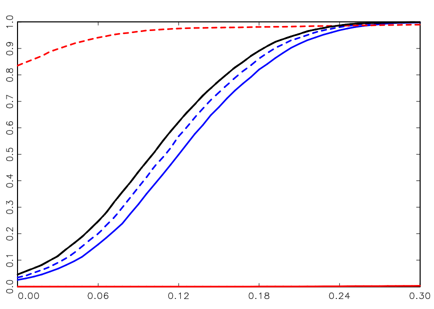
(e) $c = 50$



(f) $c = 100$



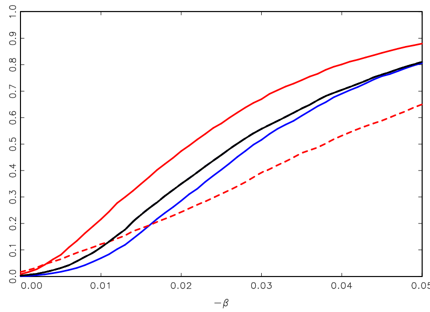
(g) $c = 250$



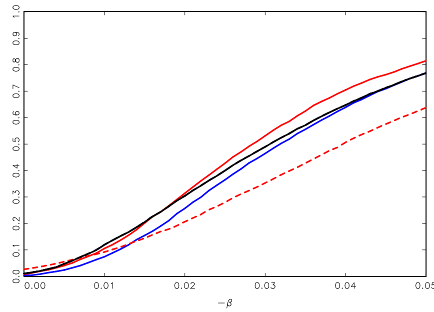
t^{GLS}: ——— , t^{OLS}: – – , Q^{GLS}: ———, Q^{OLS}: – –, U^{hyb}———, W_1^{hyb}: – –, W_2^{hyb}: –·–

Figure S.13: Finite Sample Power Left Tailed Tests - $T = 250$, $\delta = -0.95$

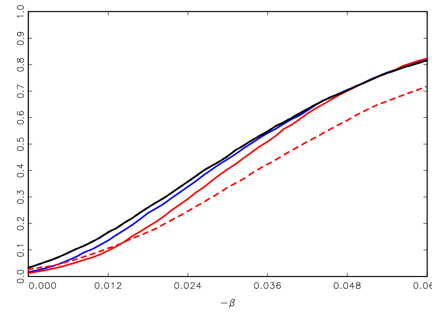
(a) $c = 0$



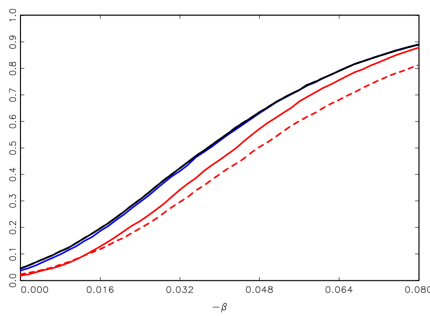
(b) $c = 2$, $\mu_\theta = 0$



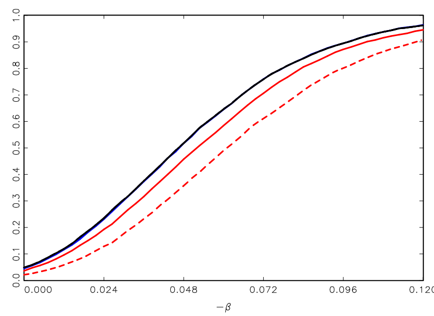
(c) $c = 5$, $\mu_\theta = 0$



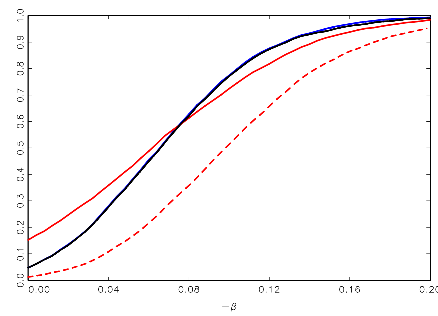
(d) $c = 10$, $\mu_\theta = 0$



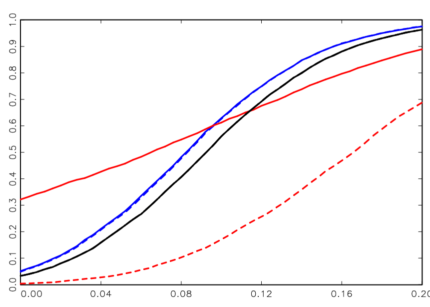
(e) $c = 20$, $\mu_\theta = 0$



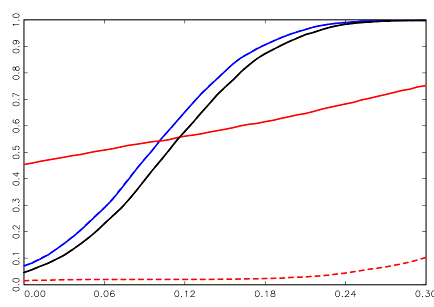
(f) $c = 50$, $\mu_\theta = 0$



(g) $c = 100$, $\mu_\theta = 0$

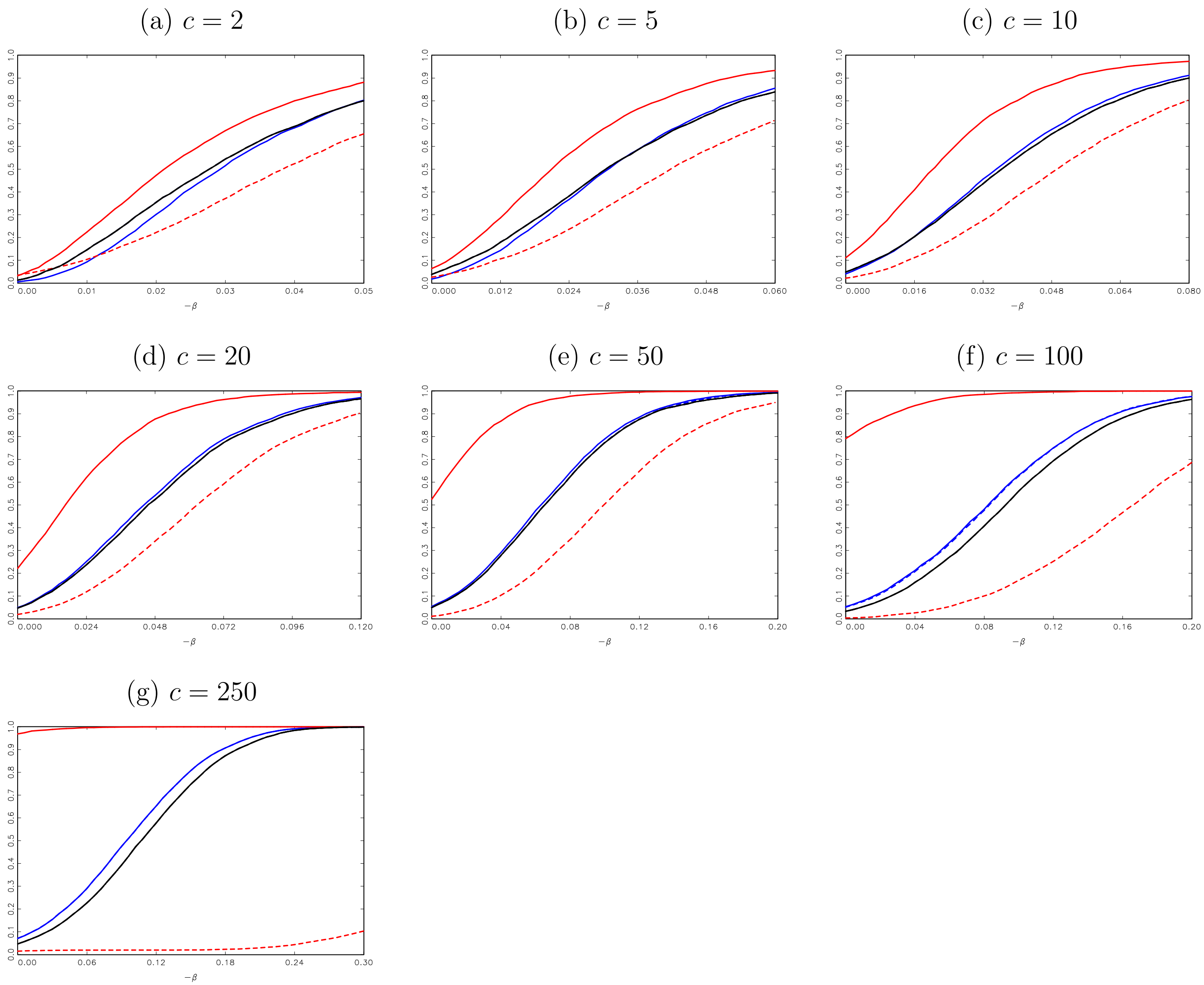


(h) $c = 250$, $\mu_\theta = 0$



t^{GLS}: ——, t^{OLS}: – –, Q^{GLS}: ——, Q^{OLS}: – –, U^{hyb}——, W_1^{hyb}: – –, W_2^{hyb}: –•–

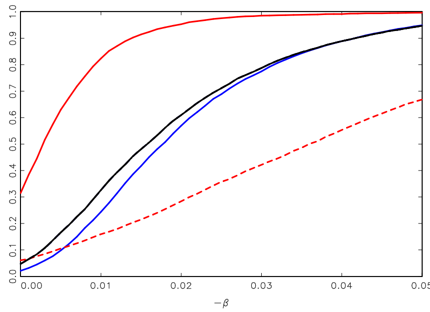
Figure S.14: Finite Sample Power Left Tailed Tests - $T = 250$, $\mu_\theta = 1$, $\delta = -0.95$



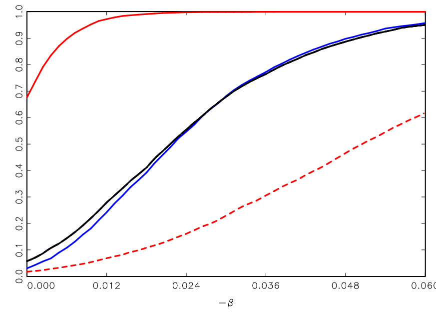
t^{GLS}: ———, t^{OLS}: – –, Q^{GLS}: ———, Q^{OLS}: – –, U^{hyb}———, W_1^{hyb}: – –, W_2^{hyb}: –·–

Figure S.15: Finite Sample Power Left Tailed Tests - $T = 250$, $\mu_\theta = 3$, $\delta = -0.95$

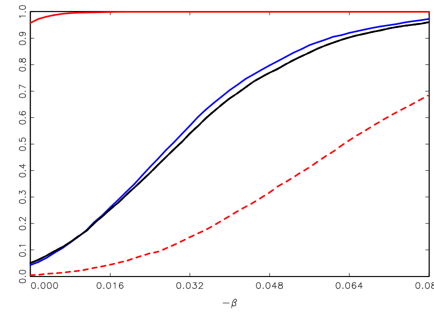
(a) $c = 2$



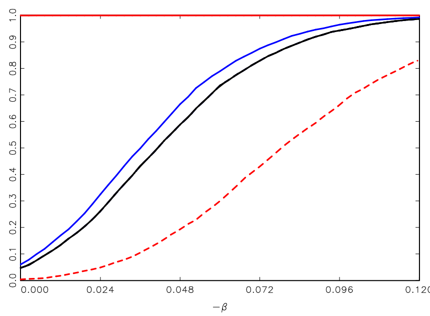
(b) $c = 5$



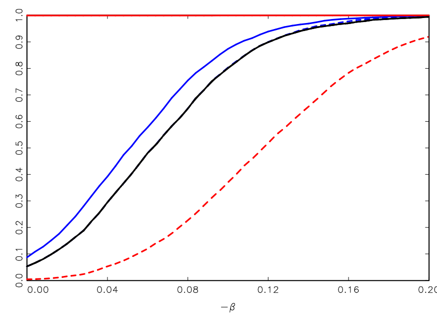
(c) $c = 10$



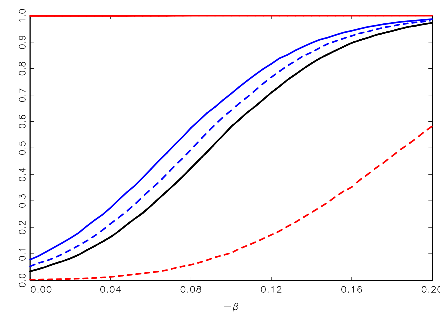
(d) $c = 20$



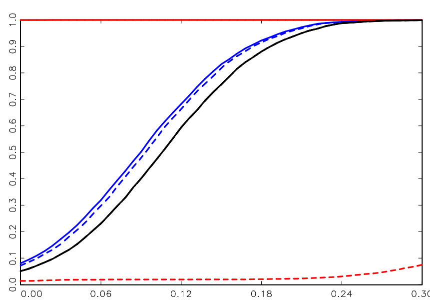
(e) $c = 50$



(f) $c = 100$

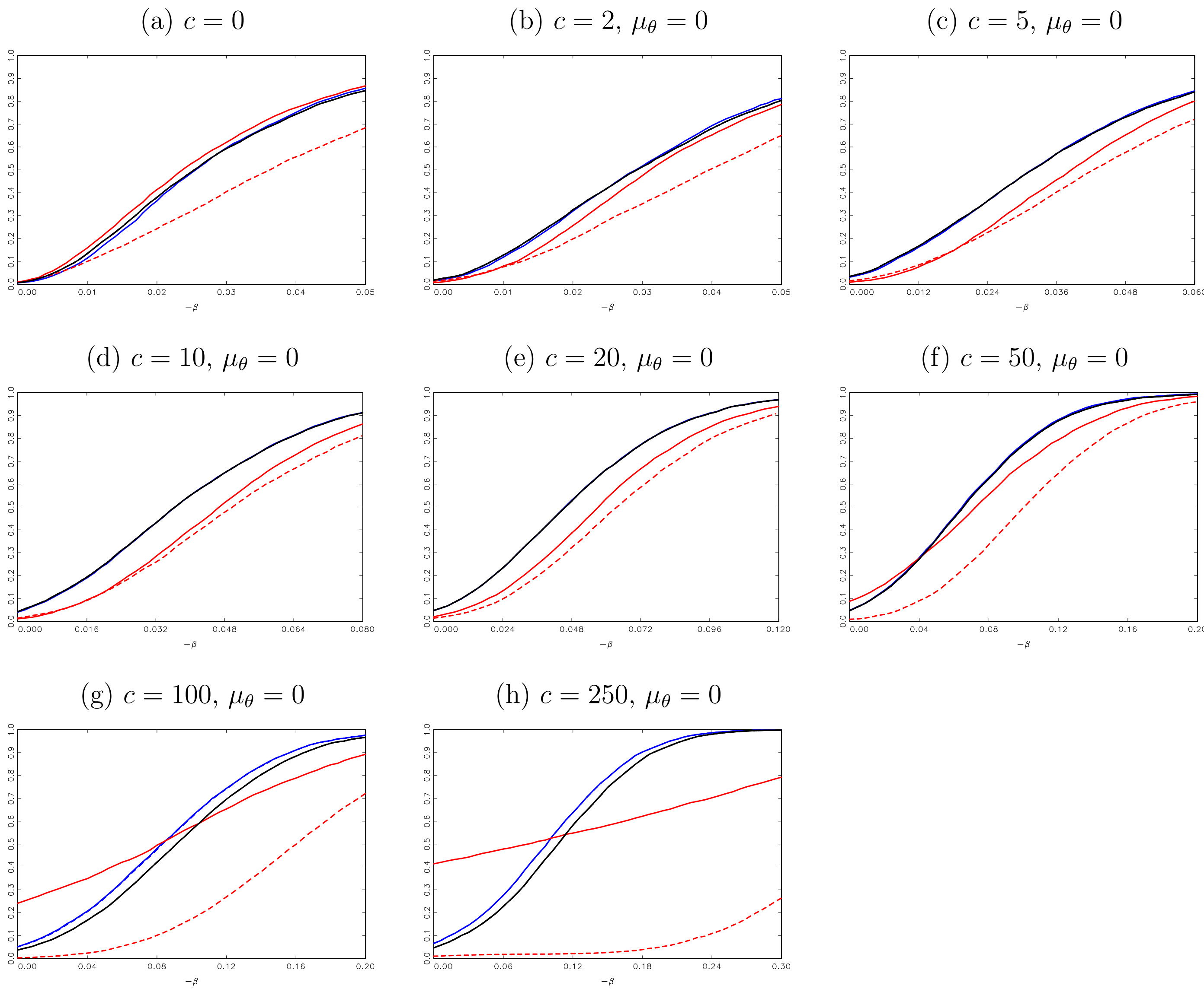


(g) $c = 250$



t^{GLS}: ——— , t^{OLS}: – – , Q^{GLS}: ———, Q^{OLS}: – –, U^{hyb}———, W_1^{hyb}: – –, W_2^{hyb}: –•–

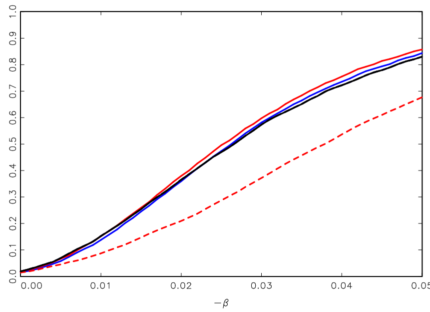
Figure S.16: Finite Sample Power Left Tailed Tests - $T = 250$, $\delta = -0.75$



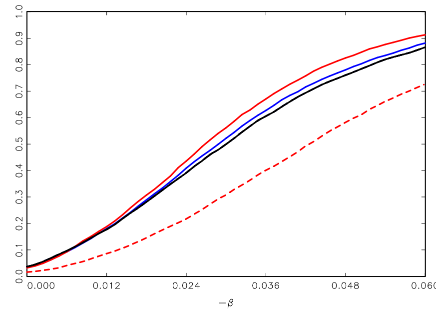
t^{GLS}: ——— , t^{OLS}: – – , Q^{GLS}: ———, Q^{OLS}: – –, U^{hyb}———, W_1^{hyb}: – –, W_2^{hyb}: –•–

Figure S.17: Finite Sample Power Left Tailed Tests - $T = 250$, $\mu_\theta = 1$, $\delta = -0.75$

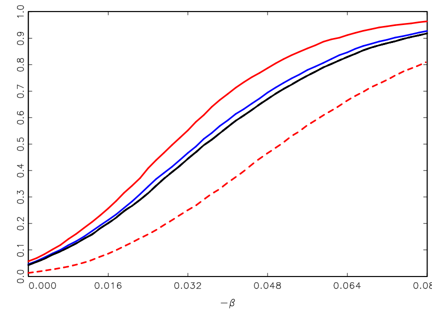
(a) $c = 2$



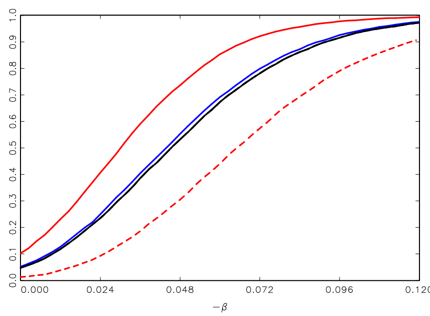
(b) $c = 5$



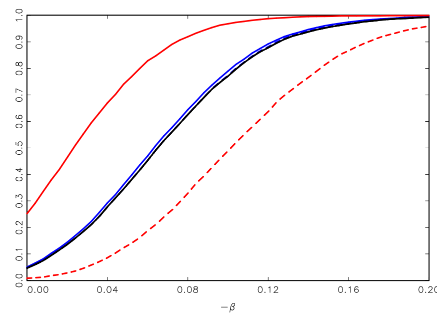
(c) $c = 10$



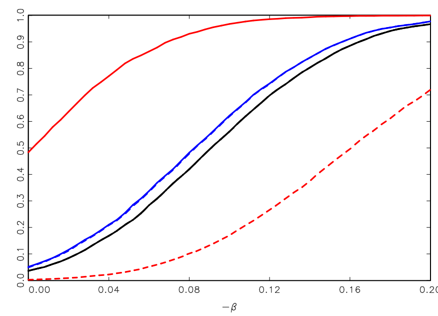
(d) $c = 20$



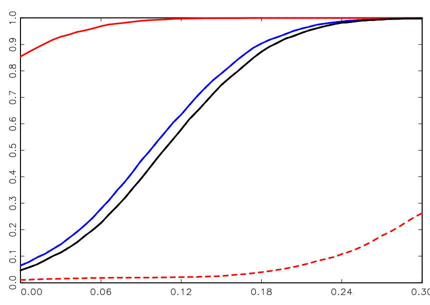
(e) $c = 50$



(f) $c = 100$

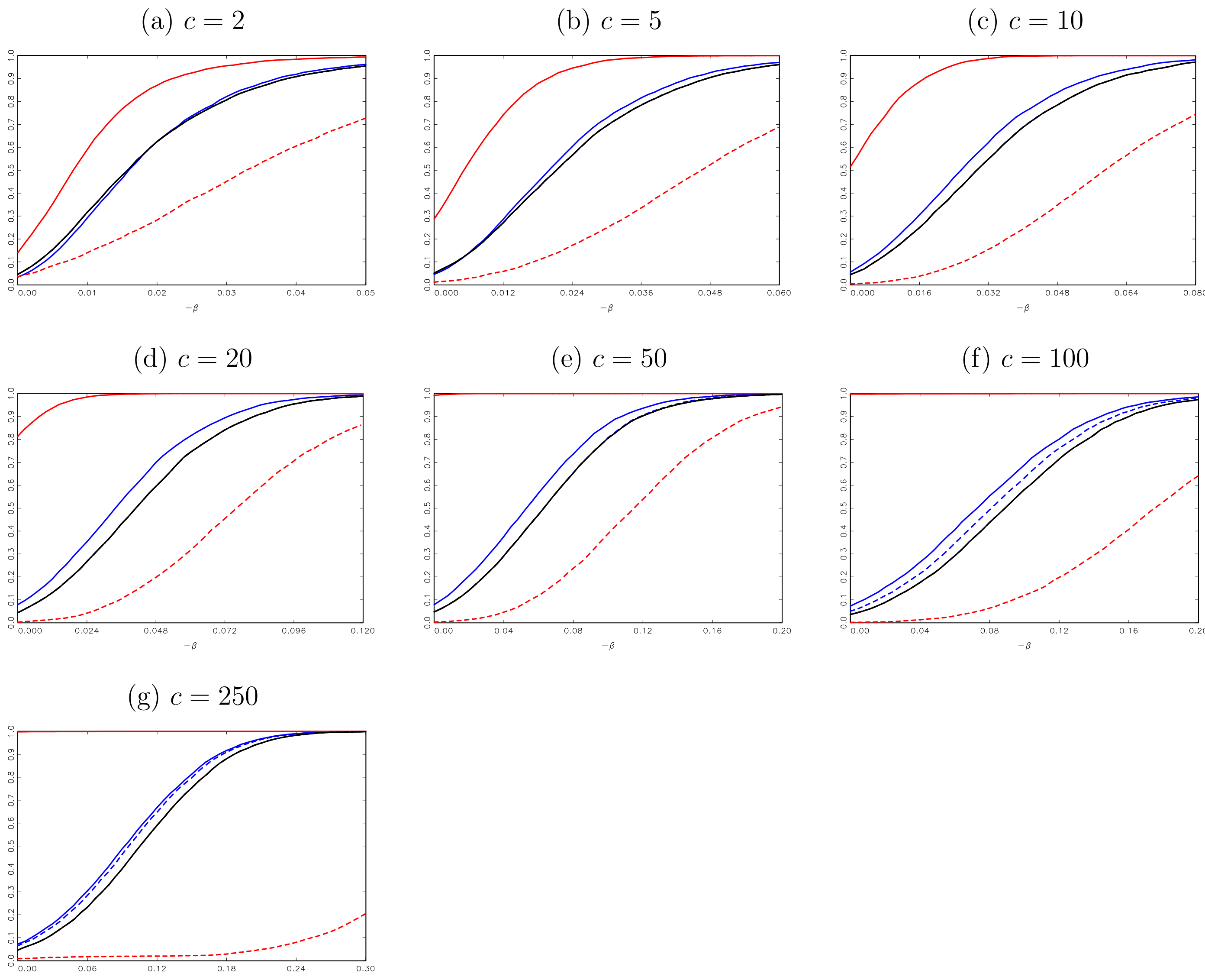


(g) $c = 250$



t^{GLS}: ——, t^{OLS}: – –, Q^{GLS}: ——, Q^{OLS}: – –, U^{hyb}——, W_1^{hyb}: – –, W_2^{hyb}: –·–

Figure S.18: Finite Sample Power Left Tailed Tests - $T = 250$, $\mu_\theta = 3$, $\delta = -0.75$



t^{GLS}: ——, t^{OLS}: – –, Q^{GLS}: ——, Q^{OLS}: – –, U^{hyb}——, W_1^{hyb}: – –, W_2^{hyb}: –·–

S.3 Empirical Application

In this section we discuss the results of the empirical application of our proposed test procedures across start dates for the examples not covered in the main paper. In all instances we consider start dates, t_s, for the predictive regressions up to and including the end of 1945 for each return/predictor pairing. The output is presented in the same format as for the examples reported in the main text.

S.3.1 S&P500 Returns 1880-2002

Figure S.19 reports the lower bound of the confidence interval for β for each test for the annual S&P500 returns data when using the earnings-price ratio as a predictor. While Q^{GLS} rejects more frequently across start dates than U^{hyb} and W_γ^{hyb}, mirroring the results in Table 2, we see that this is mainly driven by the window of start dates in the late 1910s to 1920s, for which all tests fail to reject the null of no predictability, being wider for U^{hyb} and W_γ^{hyb} than for Q^{GLS}. Of most note are the start dates t_s =1931-1934 in which $|\hat{\theta}|$ is large, leading to a run of four consecutive start dates for which Q^{GLS} fails to reject the null. The W_γ^{hyb} procedures, on the other hand, fail to reject for only two of these four start dates, and U^{hyb} fails to reject for only one, demonstrating the relative advantage of our proposed tests when the initial condition is large. While the overall rejection frequency for the t^{OLS} test is very low indeed, it also rejects for three out of these four start dates.

Figure S.20 reports results for the annual S&P500 returns series when utilising $d-p$ as a predictor. The results in this instance are rather uninteresting given that all tests reject for largely the same sequence of start dates, with the exception being W_2^{hyb} and U^{hyb} failing to reject for a small number of later start dates and the t^{OLS} test failing to reject for a vast majority of start dates. Of most note are the start dates t_s =1917,1931 which have the largest values of $|\hat{\theta}|$. While all tests fail to reject the null of no predictability for these start dates, the lower bound of the confidence interval for β is much closer to exceeding zero for U^{hyb} and W_γ^{hyb} than for Q^{GLS}.

S.3.2 CRSP Returns 1926-2002

Figure S.21 reports results for the annual CRSP return series when utilising $e-p$ as a predictor. We see that all five tests fail to reject the null of no predictability for later start dates, with the window of start dates for which U^{hyb} and W_γ^{hyb} fail to reject larger than for Q^{GLS}, and the non-rejection window for t^{OLS} test being larger still than for our proposed tests. Notably, however, there are a sequence of four start dates t_s=1931-1934 for which $|\hat{\theta}|$ is large in magnitude and for which Q^{GLS} fails to reject. For each of these four start dates, the U^{hyb}, W_2^{hyb} and t^{OLS} procedures reject the null and W_1^{hyb} only marginally fails to reject for t_s=1933.

Figure S.22 reports results for the annual CRSP return series when utilising $d-p$ as a predictor. These results show little of interest, with all tests rejecting the null of no predictability for all start dates with the exception of t_s =1926-1928,1930 for U^{hyb}, t_s =1926,1927,1930 for W_γ^{hyb} and t_s =1926-1930 for t^{OLS}. In the case of our proposed tests these non-rejections are rather marginal and the start dates are associated with fairly small values of $|\hat{\theta}|$.

Figure S.23 reports results for the quarterly CRSP return series when utilising $e-p$ as a predictor. The Q^{GLS} and W_1^{hyb} procedures reject with roughly the same frequency over the start dates in the figure, with the rejection frequency for W_2^{hyb} and U^{hyb} slightly lower overall and that for t^{OLS} very low indeed. Of most note are the start dates t_s = 1931Q4-1933Q1 where $|\hat{\theta}|$ is very large, and Q^{GLS} fails to reject for any of these start dates, whereas W_γ^{hyb}, U^{hyb} and t^{OLS} reject for each.

Figure S.24 reports results for the quarterly CRSP return series when utilising $d-p$ as a predictor. The Q^{GLS}, U^{hyb} and W_1^{hyb} tests reject with roughly the same frequency over the start dates considered in the figure, with W_2^{hyb} rejecting slightly less often and the t^{OLS} test having by far the lowest overall rejection frequency. With regards sensitivity to the initial condition, we note that the largest value of $|\hat{\theta}|$ is seen for t_s =1932Q2, and for this start date W_γ^{hyb}, U^{hyb} and t^{OLS} reject the null of no predictability, whereas Q^{GLS} fails to reject.

Figure S.25 reports results for the monthly CRSP return series when utilising $e-p$ as a predictor. While Q^{GLS} rejects with greater frequency across start dates than W_γ^{hyb}, U^{hyb} and t^{OLS} we see that this is mainly due to Q^{GLS} rejecting more often when $|\hat{\theta}|$ is small. If we instead focus on the start dates for which $|\hat{\theta}|$ is large, a much different pattern is observed. The largest values of $|\hat{\theta}|$ for this series are given by t_s =1931M12-1933M4; here, U^{hyb}, W_2^{hyb} and t^{OLS} reject for all 17 start dates, and W_1^{hyb} for 14; on the other hand,

Q^{GLS}, fails to reject for any start date in this period, further highlighting the sensitivity of this test to large initial conditions in the predictor.

Figure S.26 reports results for the monthly CRSP return series when utilising $d-p$ as a predictor. These results are less clear cut than when using $e-p$ as a predictor but we still see evidence of sensitivity of Q^{GLS} to large initial conditions, a feature not as apparent for U^{hyb} and W_{γ}^{hyb}. The largest values of $|\hat{\theta}|$ for this predictor are associated with start dates t_s = 1932M3-1932M7, with U^{hyb}, W_{γ}^{hyb} and t^{OLS} rejecting in each case while Q^{GLS} fails to do so. We also observe large values of $|\hat{\theta}|$ for t_s =1937M12-1938M4, with U^{hyb}, W_{γ}^{hyb} and t^{OLS} rejecting the null for four out of these five start dates, while Q^{GLS} again fails to do so, and from t_s = 1941M11-1942M3, with U^{hyb}, W_{γ}^{hyb} and t^{OLS} rejecting for all five dates and Q^{GLS} for none.

S.3.3 S&P500 Returns 1880-1994

Figure S.27 reports results for the annual S&P500 returns series when utilising $e-p$ as a predictor. The Q^{GLS} test fails to reject the null of no predictability for six start dates for this dataset, with all of the non-rejections associated with start dates for which $|\hat{\theta}|$ is large. In contrast, W_1^{hyb} rejects the null of no predictability for each and every possible start date, with a similar pattern observed for U^{hyb} and W_2^{hyb}, albeit with U^{hyb} and W_2^{hyb} failing to reject for t_s =1919 and W_2^{hyb} additionally failing to reject for t_s =1944. While the rejection frequency for the t^{OLS} test is higher than in other examples, it is still significantly lower than for our proposed tests.

Figure S.28 reports results for the annual S&P500 returns series when utilising $d-p$ as a predictor. The results here are rather more mixed, with Q^{GLS} rejecting for slightly more start dates than U^{hyb} and W_{γ}^{hyb}, and no start dates are identified where our proposed tests reject when Q^{GLS} fails to do so. We note, however, that for all tests the overall rejection rate across start dates is very low.

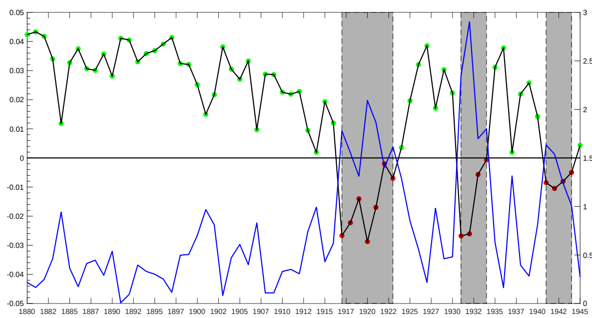
S.3.4 Annual CRSP Returns 1926-1994

Figure S.29 summarises the results for the annual CRSP returns series when utilising $e-p$ as a predictor. We see that Q^{GLS} rejects the null for a great majority of candidate start dates, but fails to reject for t_s =1931,1932 which, unsurprisingly, are the start dates for which $|\hat{\theta}|$ is large. The U^{hyb} and W_γ^{hyb} procedures, on the other hand, reject for each and every value of t_s. The t^{OLS} test fails to reject for the final two start dates where $|\hat{\theta}|$ is small.

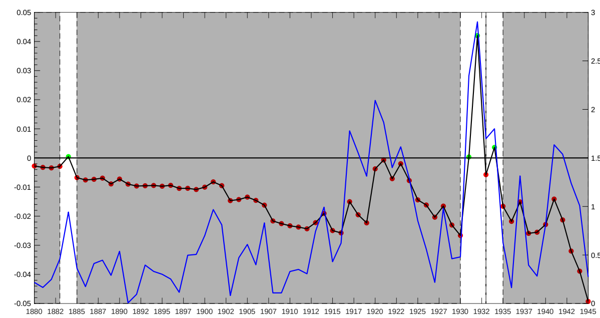
Figure S.30 reports results for the annual CRSP returns series when using $d - p$ as a predictor, with these results appearing somewhat uninteresting given that each test rejects the null of no predictability for each and every start date. We note, however, that in this instance the estimates of θ across start dates are lower overall than in most other empirical examples considered with the largest value of $|\hat{\theta}| = 2.14$, so it is unsurprising that we see less variation in rejections across start dates in this case.

Figure S.19: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Annual S&P500 1880-2002 (Predictor = $e - p$)

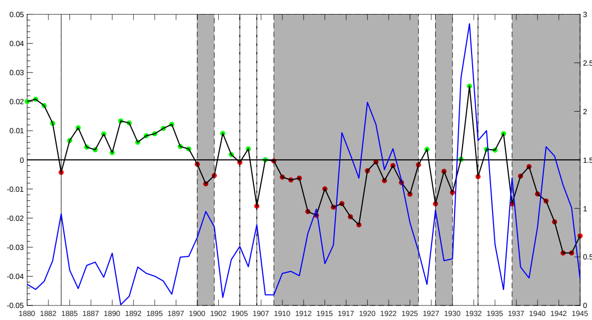
(a) Q^{GLS} (Rejection Rate 77%)



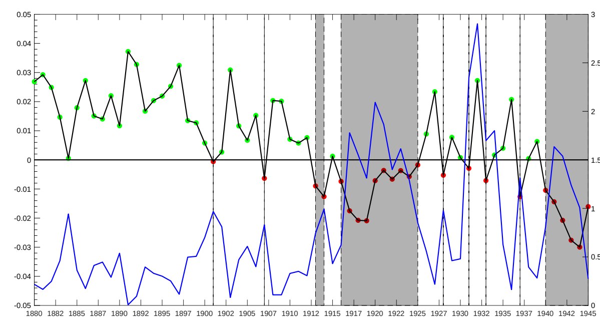
(b) t^{OLS} (Rejection Rate 6%)



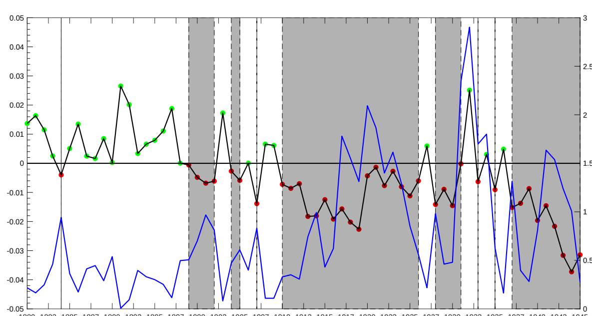
(c) U^{hyb} (Rejection Rate 44%)



(d) W_1^{hyb} (Rejection Rate 64%)



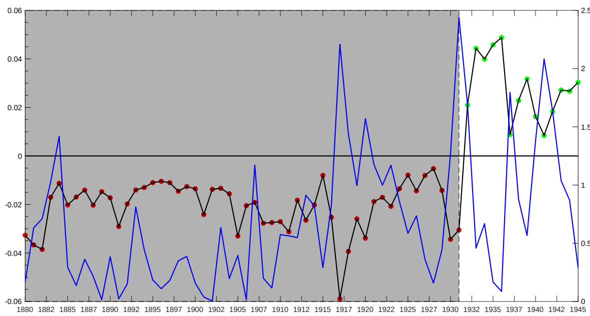
(e) W_2^{hyb} (Rejection Rate 39%)



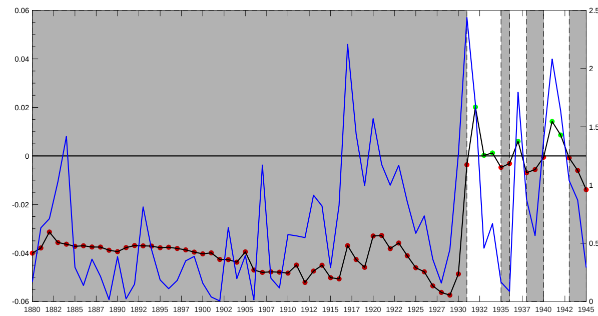
Lower Bound of CI: ——— (Left Axis) ,$|\hat{\theta}|$: ——— (Right Axis)

Figure S.20: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Annual S&P500 1880-2002 (Predictor = $d - p$)

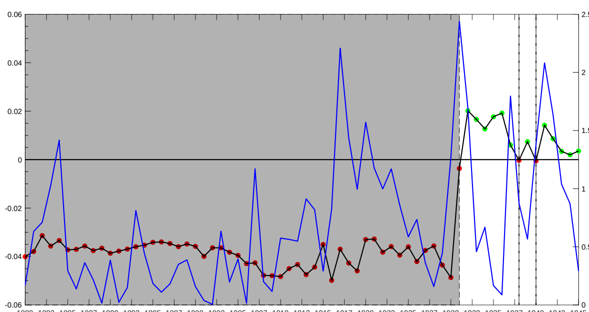
(a) Q^{GLS} (Rejection Rate 21%)



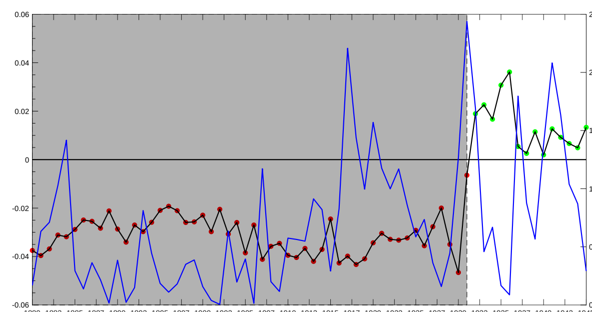
(b) t^{OLS} (Rejection Rate 9%)



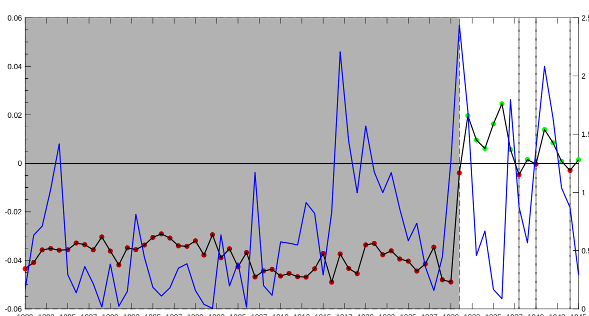
(c) U^{hyb} (Rejection Rate 18%)



(d) W_1^{hyb} (Rejection Rate 21%)



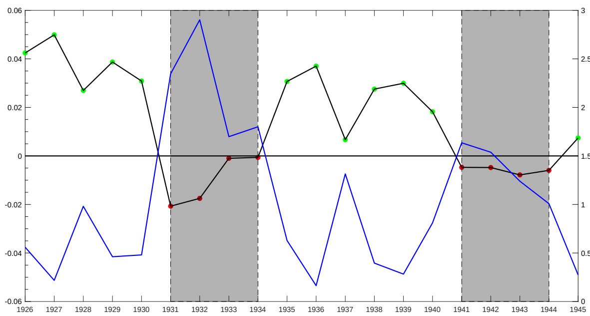
(e) W_2^{hyb} (Rejection Rate 17%)



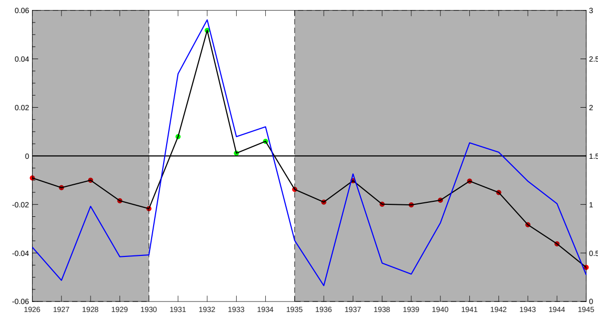
Lower Bound of CI: —— (Left Axis) ,$|\hat{\theta}|$: —— (Right Axis)

Figure S.21: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Annual CRSP 1926-2002 (Predictor = $e - p$)

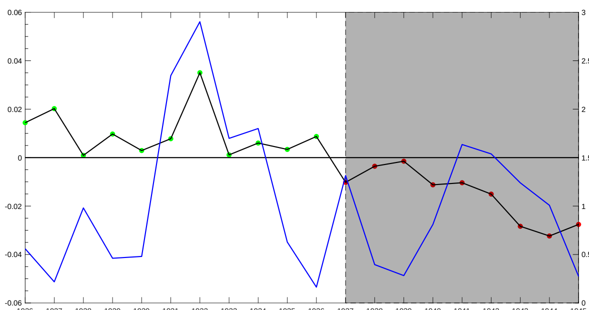
(a) Q^{GLS} (Rejection Rate 60%)



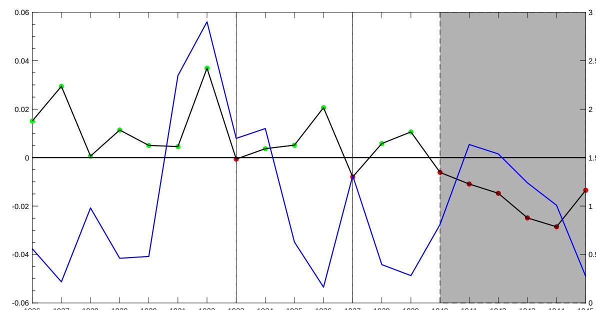
(b) t^{OLS} (Rejection Rate 20%)



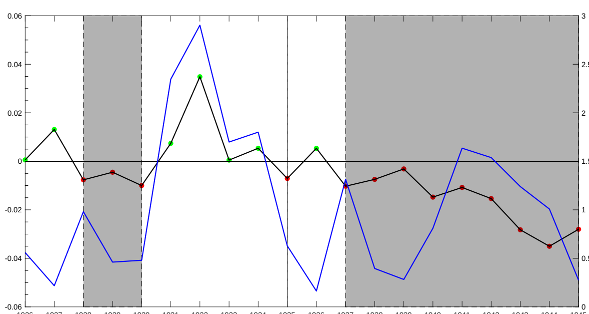
(c) U^{hyb} (Rejection Rate 55%)



(d) W_1^{hyb} (Rejection Rate 60%)



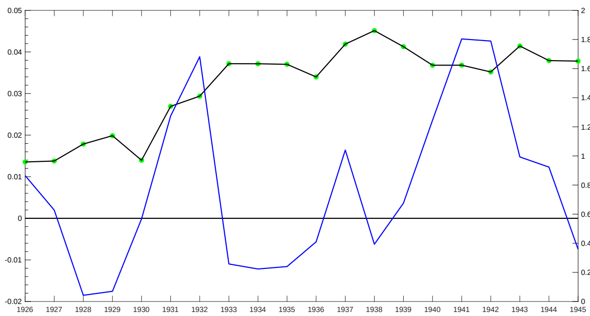
(e) W_2^{hyb} (Rejection Rate 35%)



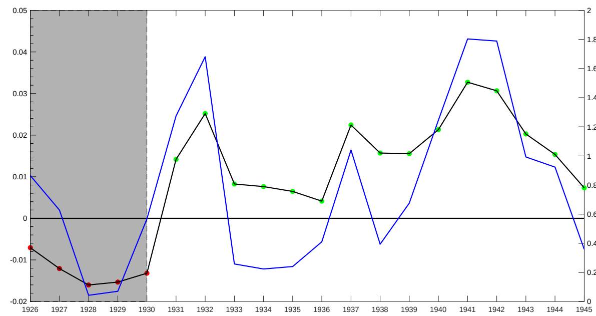
Lower Bound of CI: —— (Left Axis) ,$|\hat{\theta}|$: —— (Right Axis)

Figure S.22: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Annual CRSP 1926-2002 (Predictor = $d - p$)

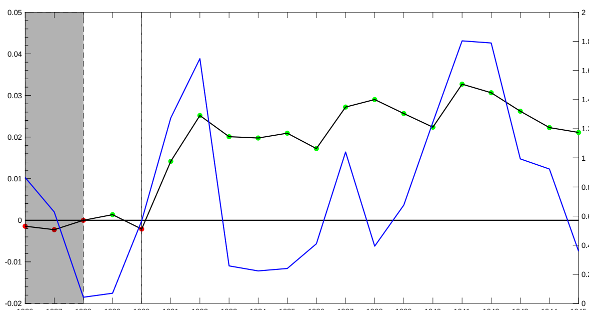
(a) Q^{GLS} (Rejection Rate 100%)



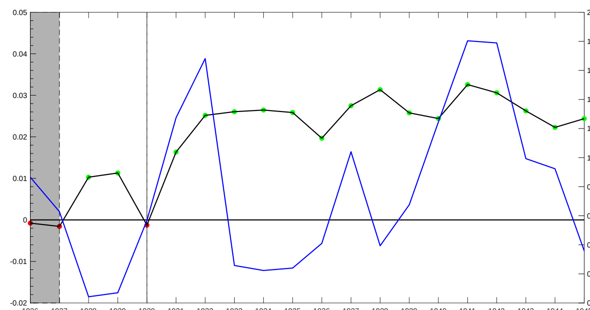
(b) t^{OLS} (Rejection Rate 75%)



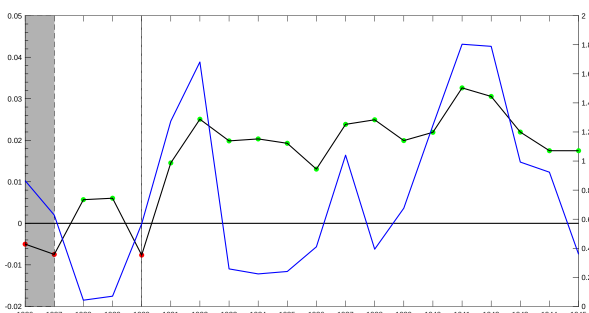
(c) U^{hyb} (Rejection Rate 80%)



(d) W_1^{hyb} (Rejection Rate 85%)



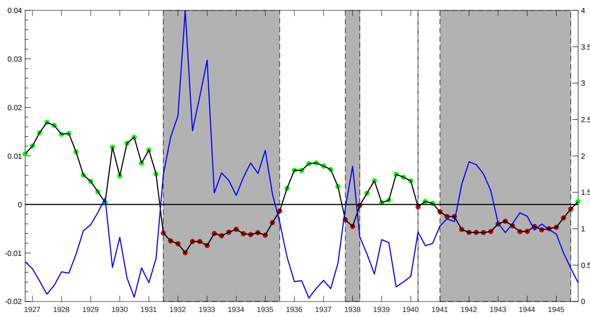
(e) W_2^{hyb} (Rejection Rate 85%)



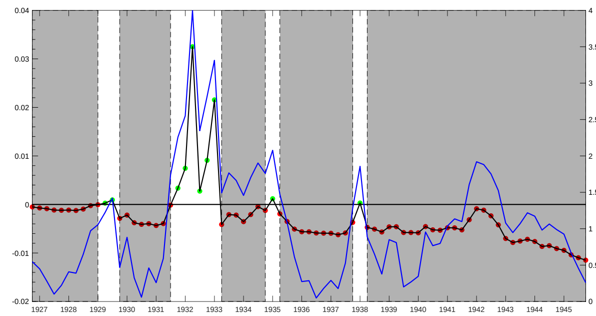
Lower Bound of CI: —— (Left Axis) ,$|\hat{\theta}|$: —— (Right Axis)

Figure S.23: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Quarterly CRSP 1926-2002 (Predictor = $e - p$)

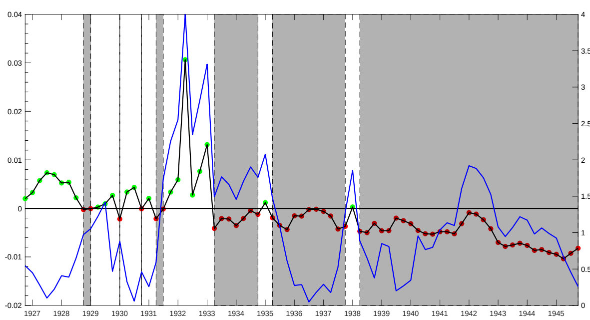
(a) Q^{GLS} (Rejection Rate 48%)



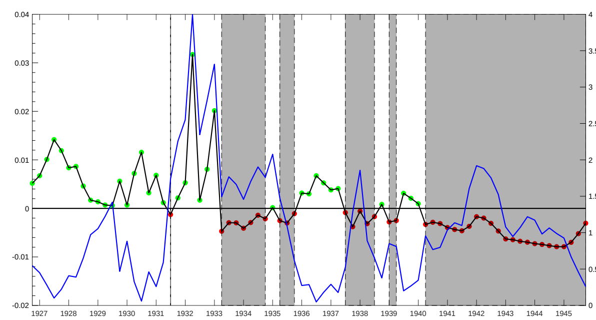
(b) t^{OLS} (Rejection Rate 13%)



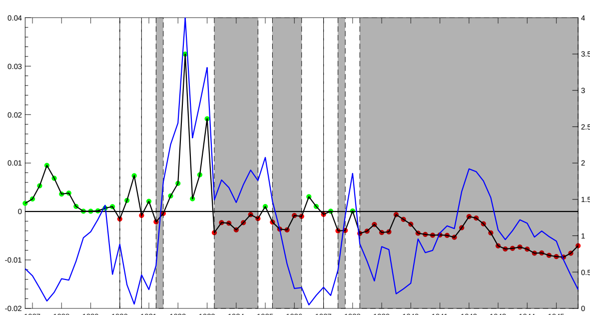
(c) U^{hyb} (Rejection Rate 29%)



(d) W_1^{hyb} (Rejection Rate 47%)



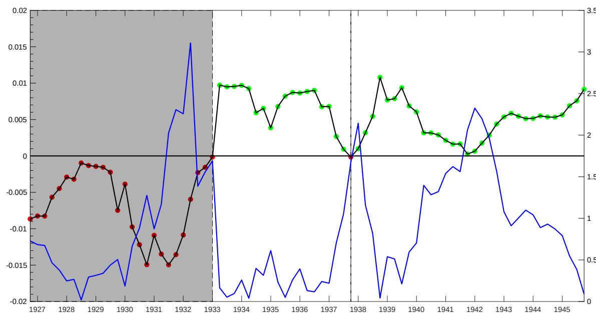
(e) W_2^{hyb} (Rejection Rate 35%)



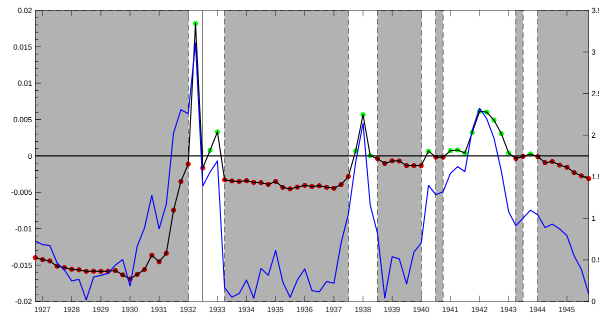
Lower Bound of CI: —— (Left Axis) ,$|\hat{\theta}|$: —— (Right Axis)

Figure S.24: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Quarterly CRSP 1926-2002 (Predictor = $d - p$)

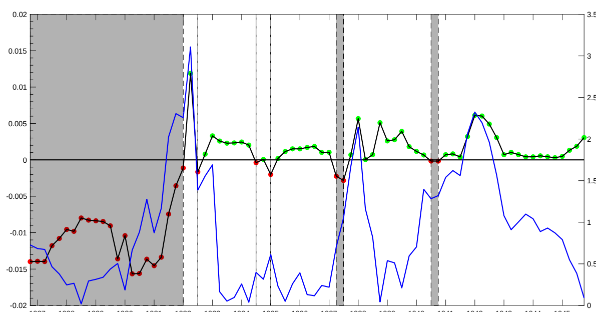
(a) Q^{GLS} (Rejection Rate 65%)



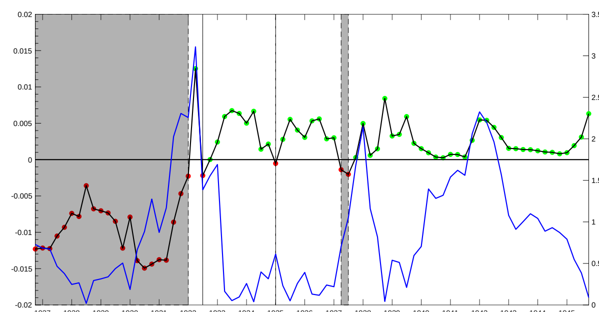
(b) t^{OLS} (Rejection Rate 22%)



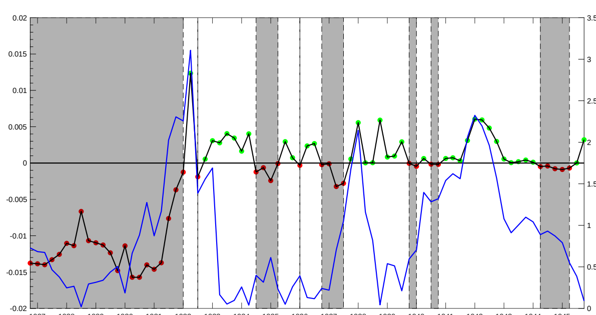
(c) U^{hyb} (Rejection Rate 62%)



(d) W_1^{hyb} (Rejection Rate 66%)



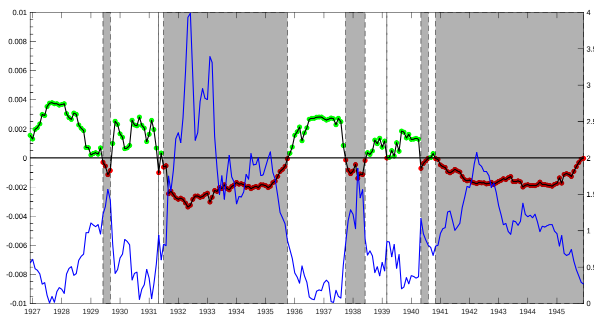
(e) W_2^{hyb} (Rejection Rate 47%)



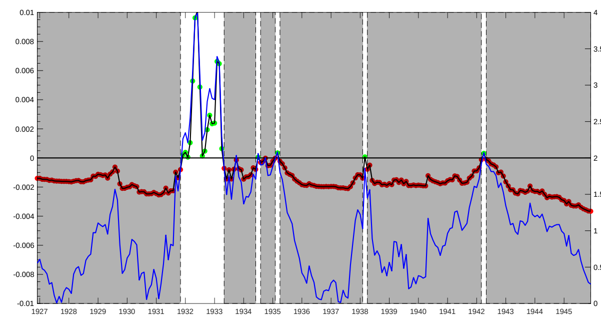
Lower Bound of CI: —— (Left Axis) ,$|\hat{\theta}|$: —— (Right Axis)

Figure S.25: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Monthly CRSP 1926-2002 (Predictor = $e - p$)

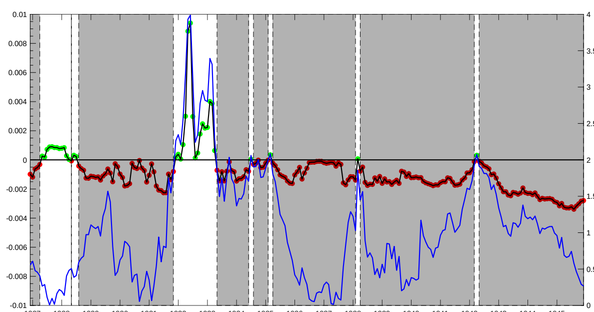
(a) Q^{GLS} (Rejection Rate 42%)



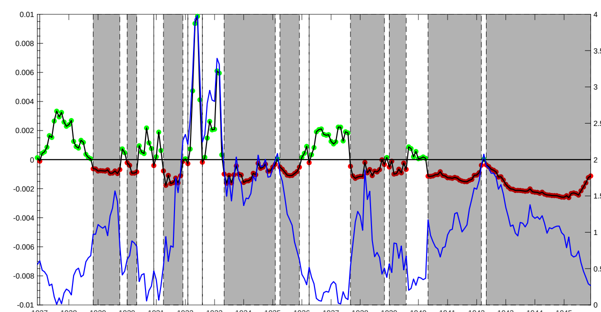
(b) t^{OLS} (Rejection Rate 9%)



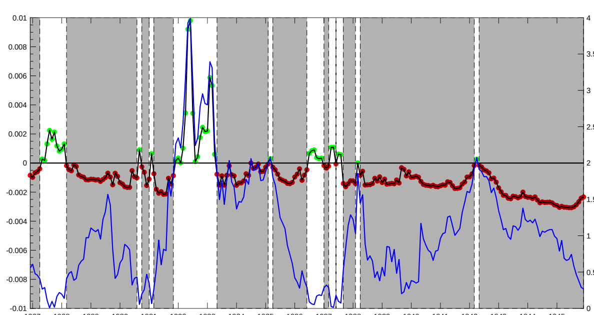
(c) U^{hyb} (Rejection Rate 15%)



(d) W_1^{hyb} (Rejection Rate 34%)



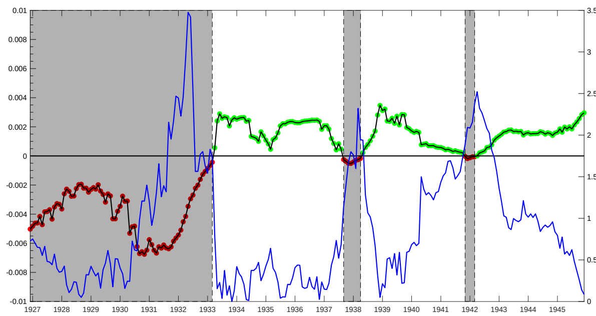
(e) W_2^{hyb} (Rejection Rate 18%)



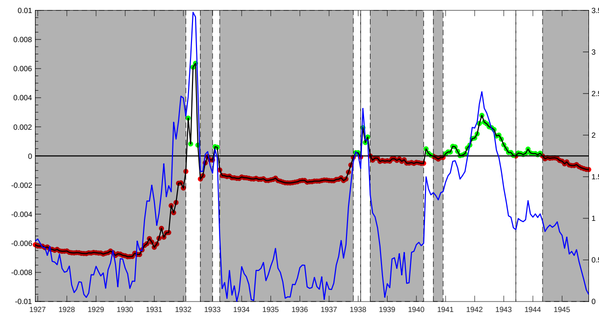
Lower Bound of CI: —— (Left Axis) ,$|\hat{\theta}|$: —— (Right Axis)

Figure S.26: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Monthly CRSP 1926-2002 (Predictor = $d - p$)

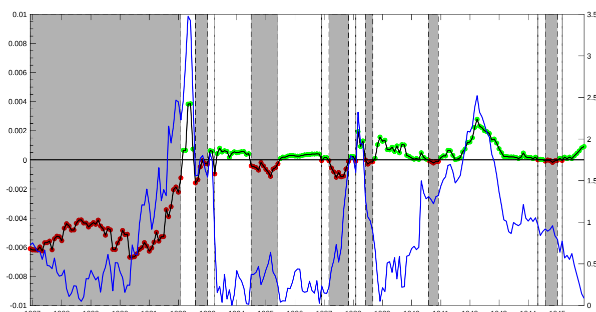
(a) Q^{GLS} (Rejection Rate 61%)



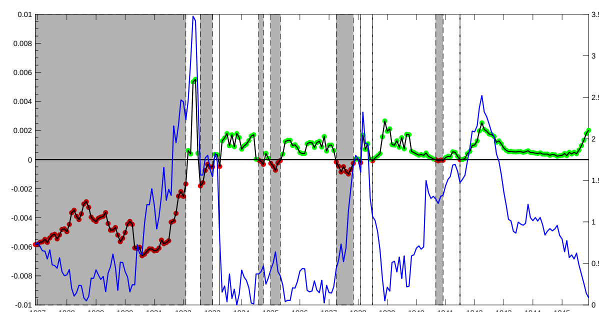
(b) t^{OLS} (Rejection Rate 24%)



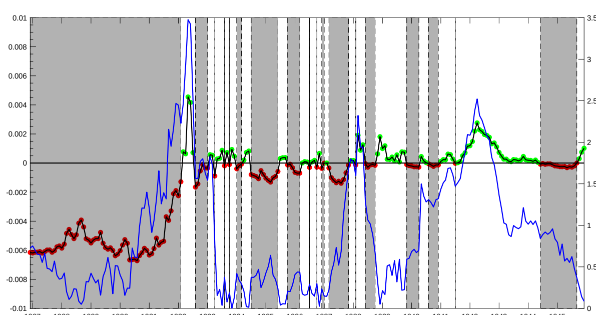
(c) U^{hyb} (Rejection Rate 52%)



(d) W_1^{hyb} (Rejection Rate 59%)



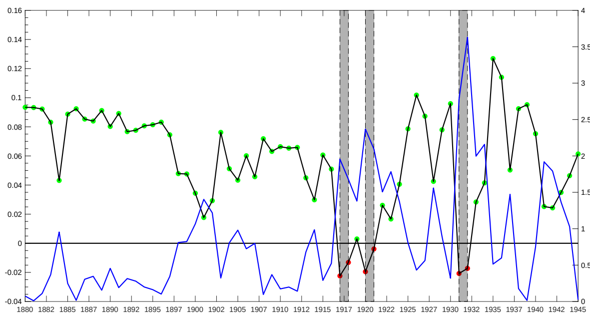
(e) W_2^{hyb} (Rejection Rate 39%)



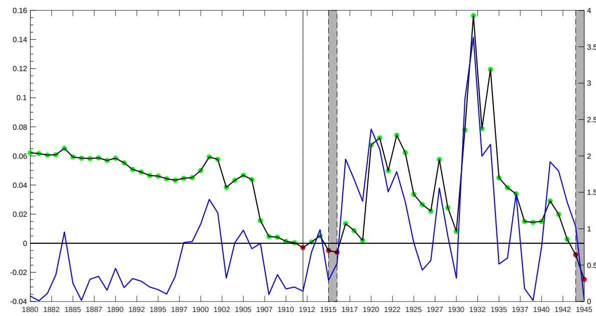
Lower Bound of CI: ——— (Left Axis) ,$|\hat{\theta}|$: ——— (Right Axis)

Figure S.27: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Annual S&P500 1880-1994 (Predictor = $e - p$)

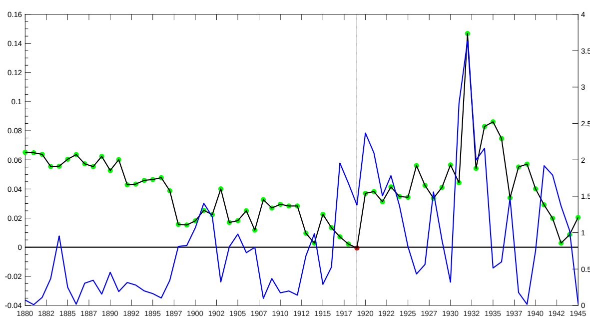
(a) Q^{GLS} (Rejection Rate 91%)



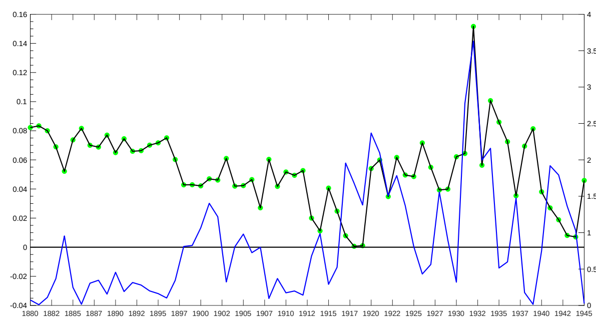
(b) t^{OLS} (Rejection Rate 92%)



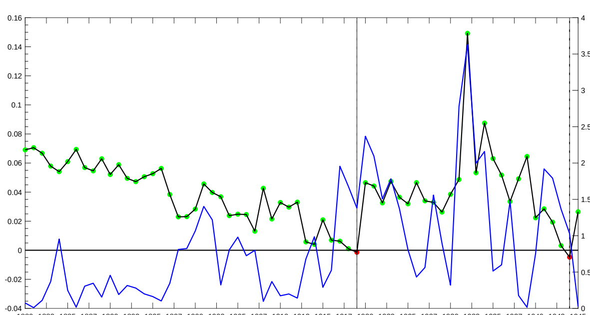
(c) U^{hyb} (Rejection Rate 99%)



(d) W_1^{hyb} (Rejection Rate 100%)



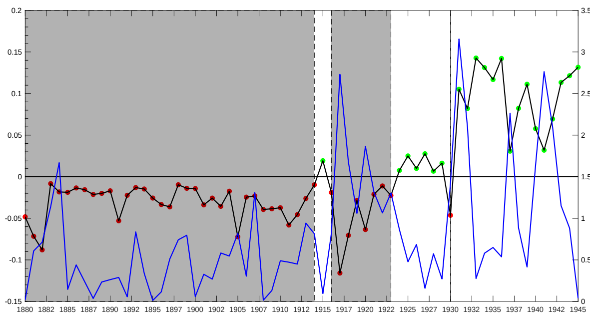
(e) W_2^{hyb} (Rejection Rate 97%)



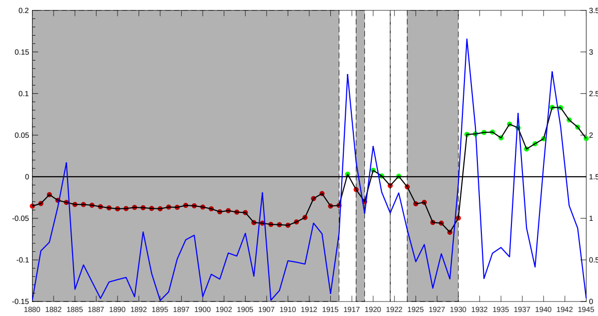
Lower Bound of CI: ——— (Left Axis) ,$|\hat{\theta}|$: ——— (Right Axis)

Figure S.28: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Annual S&P500 1880-1994 (Predictor = $d - p$)

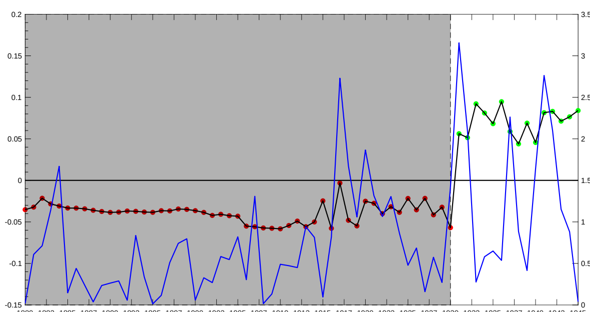
(a) Q^{GLS} (Rejection Rate 33%)



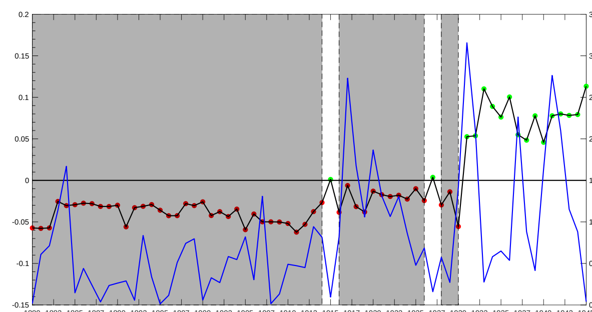
(b) t^{OLS} (Rejection Rate 29%)



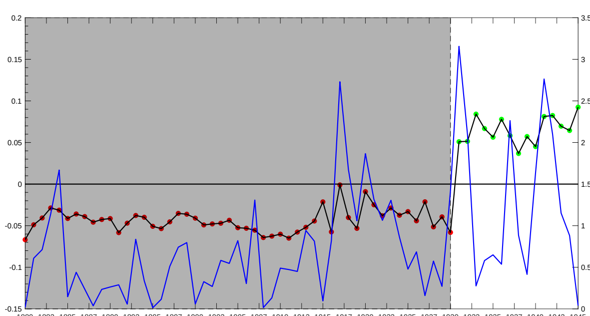
(c) U^{hyb} (Rejection Rate 23%)



(d) W_1^{hyb} (Rejection Rate 26%)



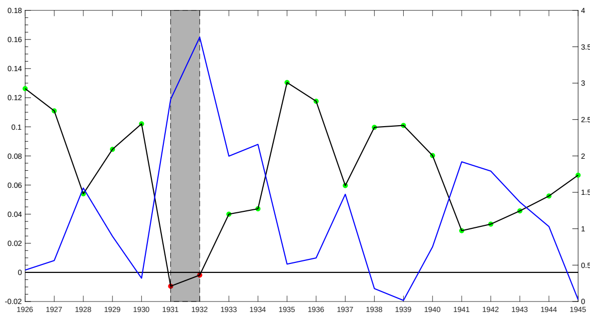
(e) W_2^{hyb} (Rejection Rate 23%)



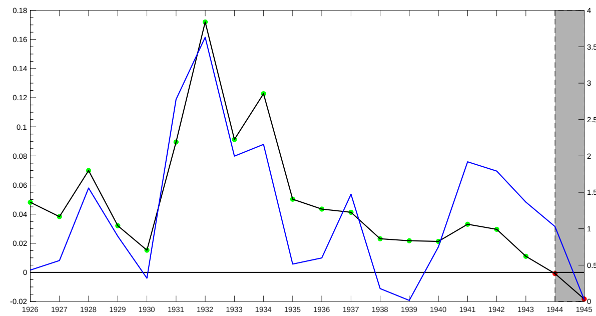
Lower Bound of CI: ——— (Left Axis) ,$|\hat{\theta}|$: ——— (Right Axis)

Figure S.29: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Annual CRSP 1926-1994 (Predictor = $e - p$)

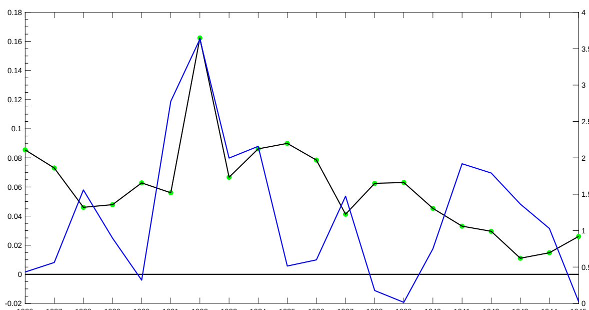
(a) Q^{GLS} (Rejection Rate 90%)



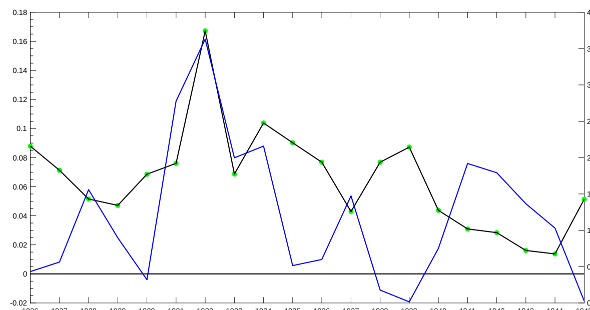
(b) t^{OLS} (Rejection Rate 90%)



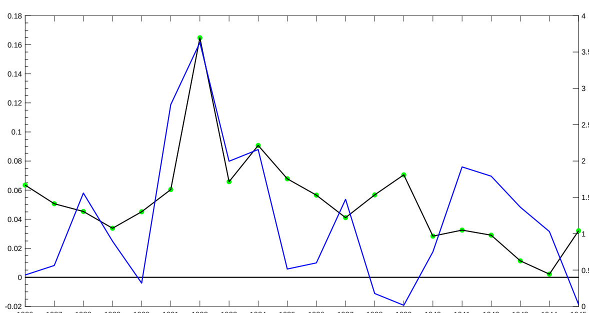
(c) U^{hyb} (Rejection Rate 100%)



(d) W_1^{hyb} (Rejection Rate 100%)



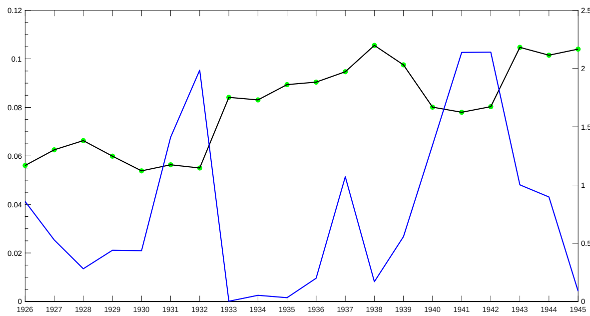
(e) W_2^{hyb} (Rejection Rate 100%)



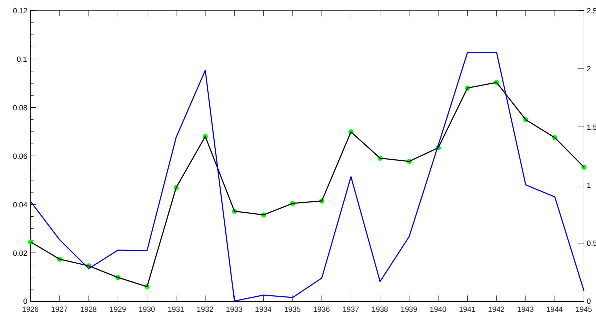
Lower Bound of CI: —— (Left Axis) ,$|\hat{\theta}|$: —— (Right Axis)

Figure S.30: Lower Bound of Confidence Interval and Estimated Magnitude of Initial Condition - Annual CRSP 1926-1994 (Predictor = $d - p$)

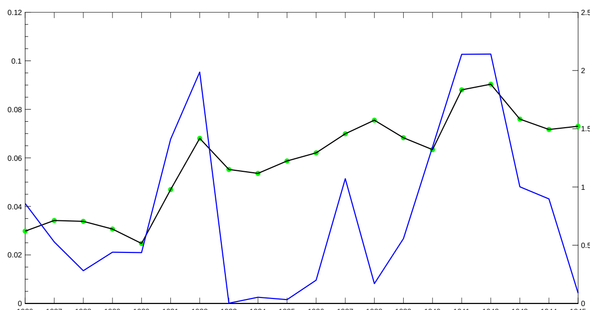
(a) Q^{GLS} (Rejection Rate 100%)



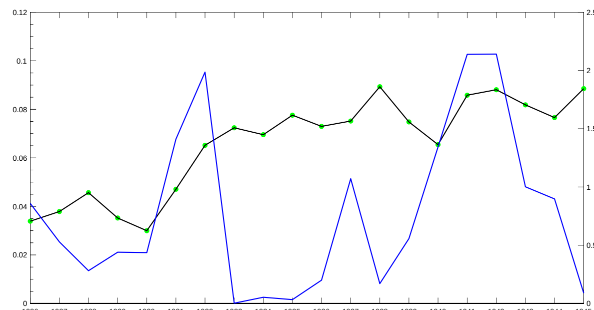
(b) t^{OLS} (Rejection Rate 100%)



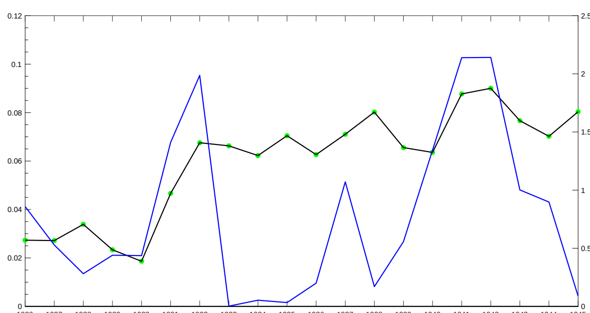
(c) U^{hyb} (Rejection Rate 100%)



(d) W_1^{hyb} (Rejection Rate 100%)



(e) W_2^{hyb} (Rejection Rate 100%)



Lower Bound of CI: ——— (Left Axis) ,$|\hat{\theta}|$: ——— (Right Axis)

S.4 Proof of Theorems

We assume, without loss of generality, that $\alpha = \mu = 0$. First we note that the estimates $\hat{\sigma}_u^2$, $\hat{\sigma}_e^2$, $\hat{\sigma}_{u,e}$, $\hat{\omega}_v^2$ and $\hat{\delta}$ are all consistent estimates of their corresponding population parameters, i.e. $\hat{\sigma}_u^2 \xrightarrow{p} \sigma_u^2$, $\hat{\sigma}_e^2 \xrightarrow{p} \sigma_e^2$, $\hat{\sigma}_{u,e} \xrightarrow{p} \sigma_{u,e}$, $\hat{\sigma}_v^2 \xrightarrow{p} \sigma_v^2$, $\hat{\omega}_v^2 \xrightarrow{p} \omega_v^2$ and $\hat{\delta} \xrightarrow{p} \delta$. In what follows, all integrals are taken over $[0, 1]$.

Proof of Theorem 2(a): As demonstrated in, for example, Harvey and Leybourne, 2005,p.111), under Assumption S.3 the following invariance principle holds,

$$T^{-1/2} x^{\mu}_{\lfloor rT \rfloor} \xrightarrow{w} \omega_v K^{\mu}_c(r)$$

where $K^{\mu}_c(r) := K_c(r) - \int_0^1 K_c(s)ds$ and $K_c(r) := \theta(e^{-rc} - 1)(2c)^{-1/2} + W_{e,c}(r)$. Using applications of the Continuous Mapping Theorem [CMT], we may then state the following key weak convergence results which will be needed to establish the large sample distributions of the t and $Q(\tilde{c})$ statistics:

$$T^{-2} \sum_{t=1}^{T} x^{\mu 2}_{t-1} \xrightarrow{w} \omega_v^2 \int K^{\mu}_c(r)^2 dr \tag{S.1}$$

$$T^{-1} \sum_{t=1}^{T} x^{\mu}_{t-1} u_t \xrightarrow{w} \sigma_u \omega_v \int K^{\mu}_c(s) dW_u(s) \tag{S.2}$$

$$T^{-1} \sum_{t=1}^{T} x^{\mu}_{t-1} v_t \xrightarrow{w} \omega_v^2 \int K^{\mu}_c(s) dW_e(s) + \frac{1}{2}(\omega_v^2 - \sigma_v^2). \tag{S.3}$$

The t-statistic for testing $\beta = 0$ under the local alternative $\beta = T^{-1}b$ can be written as

$$\begin{aligned}
t &= \frac{\hat{\beta}}{\left(\hat{\sigma}_u^2 / \sum_{t=1}^{T} x^{\mu 2}_{t-1}\right)^{1/2}} \\
&= \frac{\beta}{\left(\hat{\sigma}_u^2 / \sum_{t=1}^{T} x^{\mu 2}_{t-1}\right)^{1/2}} + \frac{\left(\sum_{t=1}^{T} x^{\mu 2}_{t-1}\right)^{-1} \left(\sum_{t=1}^{T} x^{\mu}_{t-1} u_t\right)}{\left(\hat{\sigma}_u^2 / \sum_{t=1}^{T} x^{\mu 2}_{t-1}\right)^{1/2}} \\
&= \hat{\sigma}_u^{-1} \beta \left(\sum_{t=1}^{T} x^{\mu 2}_{t-1}\right)^{1/2} + \hat{\sigma}_u^{-1} \left(\sum_{t=1}^{T} x^{\mu 2}_{t-1}\right)^{-1/2} \left(\sum_{t=1}^{T} x^{\mu}_{t-1} u_t\right) \\
&= \hat{\sigma}_u^{-1} b \left(T^{-2} \sum_{t=1}^{T} x^{\mu 2}_{t-1}\right)^{1/2} + \hat{\sigma}_u^{-1} \left(T^{-2} \sum_{t=1}^{T} x^{\mu 2}_{t-1}\right)^{-1/2} \left(T^{-1} \sum_{t=1}^{T} x^{\mu}_{t-1} u_t\right).
\end{aligned} \tag{S.4}$$

Using the results in (S.1) and (S.2) together with applications of the CMT, it therefore follows from (S.4) that

$$\begin{aligned} t &\stackrel{w}{\to} \sigma_u^{-1} b \left(\omega_v^2 \int K_c^\mu(r)^2 dr \right)^{1/2} + \sigma_u^{-1} \left(\omega_v^2 \int K_c^\mu(r)^2 dr \right)^{-1/2} \left(\sigma_u \omega_v \int K_c^\mu(s) dW_u(s) \right) \\ &= \sigma_u^{-1} \omega_v b \left(\int K_c^\mu(r)^2 dr \right)^{1/2} + \left(\int K_c^\mu(r)^2 dr \right)^{-1/2} \left(\int K_c^\mu(s) dW_u(s) \right). \end{aligned} \quad \text{(S.5)}$$

Following Cavanagh, Elliott and Stock (1995,p.1134) we can express W_u as

$$W_u = \delta W_e + (1 - \delta^2)^{1/2} \tilde{W}_u \quad \text{(S.6)}$$

where $\tilde{W}_u$ is a standard Brownian motion distributed independently of W_e. Substituting (S.6) into (S.5) we obtain that

$$\begin{aligned} t &\stackrel{w}{\to} \sigma_u^{-1} \omega_v b \left(\int K_c^\mu(r)^2 dr \right)^{1/2} + \delta \left(\int K_c^\mu(r)^2 dr \right)^{-1/2} \left(\int K_c^\mu(s) dW_e(s) \right) \\ &\quad + (1 - \delta^2)^{1/2} \left(\int K_c^\mu(r)^2 dr \right)^{-1/2} \left(\int K_c^\mu(s) d\tilde{W}_u(s) \right) \\ &= \sigma_u^{-1} \omega_v b \left(\int K_c^\mu(r)^2 dr \right)^{1/2} + \delta \left(\int K_c^\mu(r)^2 dr \right)^{-1/2} \left(\int K_c^\mu(s) dW_e(s) \right) + (1 - \delta^2)^{1/2} Z \end{aligned}$$

where Z is a standard normal random variable, thereby completing the proof. Notice that the final term above obtains as a result of $\tilde{W}_u$ being independently distributed of W_e and, therefore, also independently distributed of $K_c^\mu(r)$.

Proof of Theorem 2(b): Under the local alternative $\beta = T^{-1} b$, defining the estimate of the true value of c as $\tilde{c}$, CY show that the Q statistic for testing $\beta = 0$ can be written as

$$\begin{aligned} Q(\tilde{c}) &= \frac{b(T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2})^{1/2}}{\sigma_u (1 - \delta^2)^{1/2}} + \frac{\delta(\tilde{c} - c)(T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2})^{1/2}}{\omega_v (1 - \delta^2)^{1/2}} \\ &\quad + \frac{T^{-1} \sum_{t=1}^T x_{t-1}^\mu (u_t - \frac{\sigma_{ue}}{\sigma_e \omega_v} v_t) + \frac{1}{2} \frac{\sigma_{ue}}{\sigma_e \omega_v} (\omega_v^2 - \sigma_v^2)}{\sigma_u (1 - \delta^2)^{1/2} (T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2})^{1/2}}. \end{aligned} \quad \text{(S.7)}$$

We will derive limiting expressions for each of the three terms in the left hand side of (S.7) in turn, in each case using applications of the CMT. The first term satisfies

$$\frac{b(T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2})^{1/2}}{\sigma_u(1-\delta^2)^{1/2}} \overset{w}{\to} \frac{b\omega_v\left(\int K_c^{\mu}(r)^2 dr\right)^{1/2}}{\sigma_u(1-\delta^2)^{1/2}}. \tag{S.8}$$

The second term satisfies

$$\frac{\delta(\tilde{c}-c)(T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2})^{1/2}}{\omega_v(1-\delta^2)^{1/2}} \overset{w}{\to} \frac{\delta(\tilde{c}-c)\left(\int K_c^{\mu}(r)^2 dr\right)^{1/2}}{(1-\delta^2)^{1/2}}. \tag{S.9}$$

Turning to the final term,

$$\begin{aligned}
&\frac{T^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu}(u_t - \frac{\sigma_{ue}}{\sigma_e\omega_v}v_t) + \frac{1}{2}\frac{\sigma_{ue}}{\sigma_e\omega_v}(\omega_v^2-\sigma_v^2)}{\sigma_u(1-\delta^2)^{1/2}(T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2})^{1/2}} \\
&\quad= \frac{T^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu}u_t - \frac{\sigma_{ue}}{\sigma_e\omega_v}T^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu}v_t + \frac{1}{2}\frac{\sigma_{ue}}{\sigma_e\omega_v}(\omega_v^2-\sigma_v^2)}{\sigma_u(1-\delta^2)^{1/2}(T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2})^{1/2}} \\
&\quad\overset{w}{\to} \frac{\sigma_u\omega_v\int K_c^{\mu}(s)dW_u(s) - \frac{\sigma_{ue}}{\sigma_e\omega_v}\omega_v^2\int K_c^{\mu}(s)dW_e(s) - \frac{1}{2}\frac{\sigma_{ue}}{\sigma_e\omega_v}(\omega_v^2-\sigma_v^2) + \frac{1}{2}\frac{\sigma_{ue}}{\sigma_e\omega_v}(\omega_v^2-\sigma_v^2)}{\sigma_u(1-\delta^2)^{1/2}\omega_v\left(\int K_c^{\mu}(r)^2 dr\right)^{1/2}} \\
&\quad= \frac{\sigma_u\int K_c^{\mu}(s)dW_u(s) - \frac{\sigma_{ue}}{\sigma_e}\int K_c^{\mu}(s)dW_e(s)}{\sigma_u(1-\delta^2)^{1/2}\left(\int K_c^{\mu}(r)^2 dr\right)^{1/2}} \\
&\quad= (1-\delta^2)^{-1/2}\left(\int K_c^{\mu}(r)^2 dr\right)^{-1/2}\int K_c^{\mu}(s)dW_u(s) \\
&\qquad -\delta(1-\delta^2)^{-1/2}\left(\int K_c^{\mu}(r)^2 dr\right)^{-1/2}\int K_c^{\mu}(s)dW_e(s) \\
&\quad= (1-\delta^2)^{-1/2}\left(\int K_c^{\mu}(r)^2 dr\right)^{-1/2}\left[\int K_c^{\mu}(s)dW_u(s) - \delta\int K_c^{\mu}(s)dW_e(s)\right]
\end{aligned} \tag{S.10}$$

where we have used the weak convergence results in (S.1)-(S.3) along with applications of the CMT. Substituting (S.6) into (S.10) yields that

$$
\begin{aligned}
& (1-\delta^2)^{-1/2}\left(\int K_c^\mu(r)^2dr\right)^{-1/2} \\
& \times\left[\delta\int K_c^\mu(s)dW_e(s)+(1-\delta^2)^{1/2}\int K_c^\mu(s)d\tilde{W}_u(s)-\delta\int K_c^\mu(s)dW_e(s)\right] \\
= & (1-\delta^2)^{-1/2}\left(\int K_c^\mu(r)^2dr\right)^{-1/2}\left[(1-\delta^2)^{1/2}\int K_c^\mu(s)d\tilde{W}_u(s)\right] \\
= & \left(\int K_c^\mu(r)^2dr\right)^{-1/2}\int K_c^\mu(s)d\tilde{W}_u(s)\equiv Z\sim N(0,1). \qquad \text{(S.11)}
\end{aligned}
$$

Combining the results in (S.8), (S.9) and (S.11) we therefore have, using applications of the CMT, that

$$
Q(\tilde{c})\xrightarrow{w}\frac{b\omega_v\left(\int K_c^\mu(r)^2dr\right)^{1/2}}{\sigma_u(1-\delta^2)^{1/2}}+\frac{\delta(\tilde{c}-c)\left(\int K_c^\mu(r)^2dr\right)^{1/2}}{(1-\delta^2)^{1/2}}+Z,
$$

thereby completing the proof.

Proof of Theorem 3: The stated results follow immediately from previous results, using applications of the CMT.