Essex Finance Centre
Working Paper Series

Working Paper No 92: 06-2025

"IVX Tests for Return Predictability and the Initial Condition"

"Sam Astill, Tassos Magdalinos, and A. M. Robert Taylor"

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

IVX Tests for Return Predictability and the Initial Condition*

Sam Astill[a], Tassos Magdalinos[b] and A.M. Robert Taylor[a]

[a] Essex Business School, University of Essex

[b] University of Southampton

June 27, 2025

Abstract

We address the sensitivity of asset return predictability tests to the initial conditions of predictors. The IVX test of Kostakis *et al.* (2015, Review of Financial Studies) assumes asymptotically negligible initial conditions, which we show can result in large power losses for strongly persistent predictors. We propose a modified test that initialises the instruments at estimates of the predictors' initial conditions, enhancing robustness and detection power. Additionally, a hybrid test is introduced, combining the strengths of the original and modified tests to deliver robust performance across varying magnitude initial conditions. Empirical and simulation results demonstrate the effectiveness of these approaches in improving predictability testing.

Keywords: predictive regression; returns; initial condition; unknown regressor persistence; instrumental variable; hybrid tests.

JEL Classification: C22; C12; G14.

*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 the predictability of asset returns is an area of research that has received an increasing amount of attention in both the applied economics and finance literature. The standard approach for determining whether returns are predictable is based on a simple linear regression model with a constant and lagged (putative) predictor, which we denote x_{t-1}, with slope coefficient β. Numerous candidate predictors have been considered, with an early contribution by Fama (1981) examining the predictability of stock returns using macroeconomic variables including interest rates, industrial production, GNP and capital stock and expenditure, with other early contributions examining predictability using a variety of both macroeconomic and financial variables such as earnings and dividend price ratios, see *inter alia* Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988a,b), Fama and French (1988,1989) and Fama (1990).

While early contributions, often based inference on standard regression t-statistics on the significance of the lagged predictor, comparing to normal critical values, this approach is only valid asymptotically if the predictor is weakly persistent, or if the predictor is strongly persistent (by which we mean the predictor belongs to the local-to-unit root class of autoregressive processes) but with a zero endogeneity correlation between returns and the errors driving the predictors. Empirical evidence presented in, among others, Campbell and Yogo (2006) and Welch and Goyal (2008) suggests, however, that many, though not all, of the predictors commonly considered are highly persistent (with autoregressive roots close to unity), with a strong negative endogeneity correlation. Nelson and Kim (1993) and Stambaugh (1999) show that this causes a bias in standard ordinary least squares (OLS) coefficient estimates from the predictive regressions, while Cavanagh *et al.* (1995) show that the standard regression t-statistic does not follow a normal distribution under the null, even asymptotically, so that the tests will suffer from severe size distortions under the null.

As a result, a number of predictability tests have been developed in the literature which are designed to be asymptotically valid when the predictor is strongly persistent and endogenous. Cavanagh *et al.* (1995) suggest testing strategies that are valid for strongly persistent and endogenous predictors with a local-to-unit root autoregressive parameter given by $\rho = 1+c/n$, where n is the sample size and c a finite constant, including sup-bound, Bonferroni and Scheffe-type confidence intervals for β that are shown to have controlled size under the null. Extending this approach, Campbell and Yogo (2006) construct the (infeasible) Q test for predictability which is an optimal test when the true value of the autoregressive parameter, ρ, is known, and subsequently propose feasible Bonferroni confidence intervals for β based on an initial confidence interval for ρ obtained by inverting a unit root test statistic.

A serious drawback of the tests of Cavanagh *et al.* (1995) and Campbell and Yogo (2006) is that although they overcome the issue of dealing with a strongly persistent and endogenous predictor, they are not valid when the predictor is weakly or mildly persistent, the latter being where the predictor belongs to the mildly integrated class of processes which are such that $\rho_n = 1 + d/n^{\lambda}$, $\lambda \in (0, 1)$, where d is a negative constant. They, therefore, suffer from the same issue as the standard t-test in that they can only be used for predictors generated by a value of ρ from a subset of the parameter space of the autoregressive parameter. One could consider pre-testing using a unit root test to determine the degree of persistence of the predictor in order to choose which approach to use in practice, but while most commonly employed unit root tests will consistently reject when $\rho < 1$, they will also reject with non-zero probability when $\rho = 1 + c/n$, $c < 0$, making it impossible to reliably distinguish between a weakly stationary predictor and a strongly persistent predictor generated by a local-to-unit root autoregressive process.

An alternative strand of the literature circumvents these problems by basing predictability tests on methods of estimating the predictive regression which are robust to the properties of the regressor. Various approaches have been considered, arguably the most successful is Kostakis *et al.* (2015) [KMS] who estimate the predictive regression using the extended instrumental variable [IVX] procedure of Phillips and Magdalinos (2009). In the IVX approach each predictor in the predictive regression has an associated stochastic instrument formed by constructing a mildly integrated variable from the first differences of the predictor, initialising the instrument at zero. The IVX instrument, by construction, has lower persistence than a near-integrated variable and, as a consequence, delivers predictability statistics with asymptotically pivotal standard limiting null distributions which are valid across the entire parameter space for ρ.

The IVX test proposed by KMS is derived under the assumption that the (unobserved) *initial condition* of the predictor, denoted X_0, and defined as the deviation of the starting value of (x_t) from its unconditional mean, is asymptotically negligible in the case where the predictor is either strongly or mildly persistent; that is, $X_0 = o_p(\kappa_n^{1/2})$ where $\kappa_n = n$ for strongly persistent predictors, and $\kappa_n = n^\lambda$, $\lambda \in (0,1)$, for mildly persistent predictors.[1] This is a strong assumption to make in practice, and it is arguably of considerably more empirical relevance to allow the initial condition of the predictor to have the same asymptotic order of magnitude as the rest of the sample data in this scenario; that is, $X_0 = O_p(\kappa_n^{1/2})$. The same assumption of an $o_p(n^{1/2})$ initial condition for strongly persistent predictors is made by Campbell and Yogo (2006) in the construction of their Bonferroni Q test, and it was shown by Astill *et al.* (2024) that when $X_0 = O_p(n^{1/2})$ the Bonferroni Q test suffers from severe asymptotic undersize (oversize) when testing in the right (left) tail when the innovations to the predictor and returns are negatively correlated. In contrast, we will show that, for a strongly or mildly persistent predictor with an $O_p(\kappa_n^{1/2})$ initial condition, the IVX test of KMS continues to admit a standard normal limiting null distribution. However, in this scenario we will show, using Monte Carlo simulations, that convergence to the limiting distribution is very slow, with the IVX test displaying severe finite sample undersize and consequent, potentially catastrophic, loss of power to detect a genuine predictor.

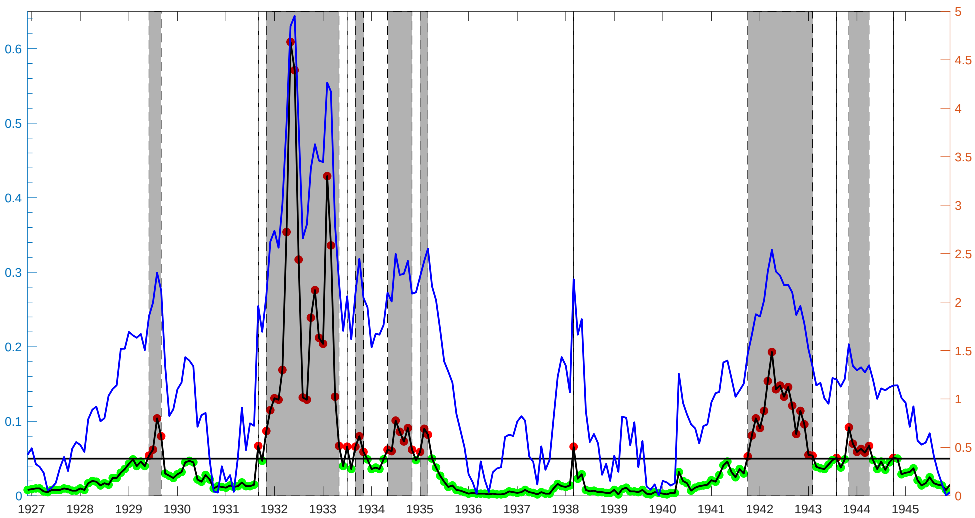
To better motivate this paper, and to demonstrate the impact of the magnitude of the initial condition of the predictor on the IVX test, we now discuss part of the results from our empirical study in Section 6 where we perform the IVX test recursively across

[1] For weakly persistent predictors, KMS argue that X_0 can be allowed to be of $o_p(n^{1/2})$ but, as we will show, this claim is incorrect and X_0 must be of $O_p(1)$ when the IVX instrument is initialised at zero; cf. Remark 2.7.

sample start dates for one of the return/predictor pairings analysed by Campbell and Yogo (2006). Specifically, we examine 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 Campbell and Yogo (2006) ($n = 817$) using a bootstrap implementation of the IVX t-statistic that we outline in detail in Section 3. For the full sample of data, such that the initial value of the predictor is its outcome at time 1926M12, we perform a test for predictability using the IVX t-statistic and compute an estimate of the magnitude of the initial condition, denoted $|\hat{\alpha}|$ and defined in (25) below, which estimates how many standard deviations the initial condition of the predictor lies from its mean. We find that the p-value of the IVX test is 0.008, giving a strong rejection of the null hypothesis of no predictability, while the value of $|\hat{\alpha}| = 0.42$ implies that the initial condition, when using the full available sample of data, is relatively small. We then repeat this exercise, but instead run the IVX test on data from $t = t_s, ..., n$, such that t_s is the time point associated with the initial value of the predictor in the regression, sequentially across the start dates t_s =1927M1,...,1945M12. The results of this exercise are summarised in Figure 1. The red/green highlighted line plots, for each start date t_s, the p value of the IVX test, with a p-value below 0.05 signalling a rejection of the null hypothesis of no predictability at the 5% significance level (green highlights) and a p-value above 0.05 signalling non-rejection of the null (red highlights). The blue line plots the estimated value of the initial condition of the predictor, $|\hat{\alpha}|$, and the grey shaded regions further highlight those start dates t_s for which the IVX test fails to reject the null of no predictability at the 5% level.

It is apparent from Figure 1 that while the IVX test rejects the null of no predictability in favour of the alternative of positive predictability for a majority (72%) of sample start dates, there are a substantial number of sample start dates for which the IVX test fails to find evidence of predictability. Further examining Figure 1 we see that there is a clear relationship between the p-value of the IVX test and the estimated magnitude of the initial condition,

Figure 1: p-value of right-tailed IVX t-test and $|\hat{\alpha}|$.



p-value: —— (Left Axis), $|\hat{\alpha}|$: —— (Right Axis)

with the sample start dates for which the IVX test finds no evidence of predictability corresponding to instances where the magnitude of the initial condition is estimated to be large. This empirical example gives an indication that the magnitude of the initial condition of the predictor has the potential to have a substantial impact on the outcome of the IVX test, and we will demonstrate this further through Monte Carlo simulation later in the paper.

The sensitivity of the IVX test to the initial condition motivates us to develop a simple modification to the IVX test that is designed to show greater robustness to the magnitude of the initial condition. This modification simply entails changing the initialisation of the IVX instrument used in the test. KMS initialise the instrument at zero. Our modification initializes the instrument at an estimate of the (unobserved) initial condition of the predictor, obtained as the observed starting value of the predictor minus its sample mean. The resulting instrument is consequently initiated at an approximation to the initial condition of the underlying predictor variable. We show that the resulting modified version of KMS's IVX test still retains a limiting standard normal null distribution. Crucially, however, numerical simulations show that its finite sample null distributions lie much closer to this limiting null distribution than do those of the original IVX test of KMS. As a result the test has finite sample size which is much closer to the nominal level and does not suffer the potentially very large power losses seen in the original IVX test of KMS when the initial condition is large. However, where the initial condition is small and the predictor is less persistent than a pure unit root series, the modified IVX test is not as powerful as the original IVX test. We therefore propose a simple hybrid test which combines information from both the original IVX statistic and the modified IVX statistic to try and capture the superior power properties of the better performing test across both large and asymptotically negligible initial conditions.

The remainder of the paper is organised as follows. The predictive regression model and assumptions under which we will work are detailed in Section 2. In Section 3 we outline the IVX test of KMS together with our modified version of this test and hybrid test which combines information from the two. The asymptotic behaviour of these statistics for both asymptotically negligible and non-negligible initial conditions are provided in Section 4. Section 5 reports results from a Monte Carlo simulation study examining the finite sample properties of the tests. Section 6 reports an empirical application of our proposed test procedures to two of the key predictors contained in the Campbell and Yogo (2006) dataset. Section 7 concludes. A supplementary appendix contains proofs of the large sample results given in Section 4, together with additional material relating to the Monte Carlo simulation exercise.

2 The Predictive Regression Model and Assumptions

Consider the following predictive regression model

$$y_t = \mu_y + \beta x_{t-1} + \varepsilon_t, \qquad t = 1, ..., n \tag{1}$$

where y_t denotes the (excess) return on a given asset in period t, and x_{t-1} denotes a putative predictor observed at time $t-1$. To aid exposition we focus attention on the case where x_t is a single (putative) predictor. The results in here generalise straightforwardly to the case where x_t is a vector of predictors; see the discussion in Remark 3.2 below. Our interest in is on testing the null hypothesis $H_0 : \beta = 0$ in (1) such that x_{t-1} is not a predictor for returns, against either one-sided ($\beta > 0$ or $\beta < 0$) or two-sided ($\beta \neq 0$) alternatives, whereby x_{t-1} is a significant predictor for returns.

The data generating process [DGP] for the candidate predictor variable, x_t, is assumed to satisfy

$$x_t = \mu + X_t \tag{2}$$

$$X_t = \rho_n X_{t-1} + u_t \tag{3}$$

for $t \geq 1$. Formal conditions will be placed on the *initial condition* of $(x_t)_{t\in\mathbb{N}}$, defined as $X_0 := X_0(n)$, in Assumption 2 below. Notice that the initial condition is defined to be the deviation of x_0 from μ, the unconditional mean of $(x_t)_{t\in\mathbb{N}}$. The (possibly sample-size-dependent) autoregressive root, ρ_n, and the innovation sequences, $(\varepsilon_t)_{t\in\mathbb{N}}$ and $(u_t)_{t\in\mathbb{N}}$, in (1) and (3), respectively, are assumed to satisfy the following set of conditions:

Assumption 1. *(a)* $\rho_n \to \rho \in (-1, 1]$ *and* $c := \lim_{n\to\infty} n(\rho_n - 1)$ *exists in* $\mathbb{R} \cup \{-\infty\}$.

(b) The innovation sequence $(u_t)_{t\in\mathbb{N}}$ *in (3) is a stationary linear process of the form*

$$u_t = \sum_{j=0}^{\infty} c_j e_{t-j}$$

where $(c_j)_{j\geq 0}$ *is a sequence of constants satisfying* $\sum_{j=0}^{\infty} |c_j| < \infty$, $\sum_{j=0}^{\infty} j c_j^2 < \infty$, $c_0 = 1$ *and* $C(1) := \sum_{j=0}^{\infty} c_j \neq 0$ *. Given a filtration* $(\mathcal{F}_t)_{t\in\mathbb{Z}}$*, the sequence* $v_t := (\varepsilon_t, e_t)'$ *is an* $\mathcal{F}_t$*-martingale difference sequence satisfying one of the following assumptions:*

(i) $\mathbb{E}_{\mathcal{F}_{t-1}}(v_t v_t') = \Sigma_v > 0$ *a.s. for all* t *and* $\left(\|v_t\|^2\right)_{t\in\mathbb{Z}}$ *is a uniformly integrable sequence.*

(ii) Denoting $\sigma_t^2 := \mathbb{E}_{\mathcal{F}_{t-1}}\varepsilon_t^2$, $\left(e_t, \mathbb{E}_{\mathcal{F}_{t-1}} e_t^2\right)_{t\in\mathbb{Z}}$ *and* $(\varepsilon_t, \sigma_t^2)_{t\in\mathbb{N}}$ *are strictly stationary with* $\mathbb{E}e_1^4 < \infty$, $\mathbb{E}\varepsilon_1^4 < \infty$ *and* $\{\sigma_t^2 - \mathbb{E}\varepsilon_1^2 : t \geq 1\}$ *is a mixingale sequence:*

$$\left\|\mathbb{E}_{\mathcal{F}_{t-1-m}}\left(\sigma_t^2 - \mathbb{E}\varepsilon_1^2\right)\right\|_{L_2} \leq b\psi_m \quad \text{for all } \ t, m \geq 1 \tag{4}$$

for some $b > 0$ *and a non-negative sequence* $(\psi_m)_{m\in\mathbb{N}}$ *satisfying* $\psi_m \to 0$.

Remark 2.1. The condition in Assumption 1(a) covers the family of autoregressive processes considered in KMS with strong, mild and weak persistence all covered. The local-to-unit root form of strong persistence occurs when $c \in \mathbb{R}$, in which case $\rho_n \sim 1 + c/n$, with the case of an exact unit root corresponding to $c = 0$. Weak persistence and mild persistence obtain when $c = -\infty$: in the former case $\rho_n \to \rho \in (-1, 1)$, while in the latter case $\rho_n \sim 1 + b/k_n$ for some $b < 0$ and a sequence $(k_n)_{n \in \mathbb{N}}$ satisfying $k_n \to \infty$ and $k_n/n \to 0$. Notice that c is allowed to take positive values in the local-to-unit root case, so convergence of ρ_n to unity may obtain from both sides of unity under strong persistence. $\diamondsuit$

Remark 2.2. The regularity conditions placed on the innovation processes (ε_t) and (u_t) in (1) and (3), respectively, by Assumption 1(b) are standard in the predictive regression literature. In particular, the innovations (ε_t) of the model are required to have the martingale difference property whereas the autoregressive innovations (u_t) may exhibit autocorrelation in the form of a short memory (weakly dependent) linear process driven by martingale difference primitive innovations, (e_t). The requirement that $C(1) \neq 0$ rules out the possibility of a moving average unit root at the long run frequency in (u_t). The moment conditions imposed on the martingale difference $v_t := (\varepsilon_t, e_t)'$, depend on the properties of its conditional variance. Under conditional homoskedasticity, only uniform integrability of $\left(\|v_t\|^2\right)$ is required, a minimal assumption for the validity of a central limit theorem on (v_t). Under conditional heteroskedasticity, (v_t) is required to have finite fourth moments and the centred conditional variance of the model's innovations (ε_t) to satisfy the mixingale property. It is well known (e.g. Example 1 of Arvanitis and Magdalinos, 2018) that the mixingale property (4) is satisfied by a stationary $ARCH(\infty)$ process: $\varepsilon_t = \eta_t \sigma_t$, where (η_t) is an IID $(0, 1)$ sequence and

$$\sigma_t^2 = \varpi + \textstyle\sum_{i=1}^{\infty} \alpha_i u_{t-i}^2, \quad \alpha_i \geq 0, \ \varpi > 0, \ \sum_{i=1}^{\infty} \alpha_i < 1 \tag{5}$$

with $\mathcal{F}_t$ the natural filtration of η_t. As a result, (4) is also satisfied by any stationary finite order GARCH process with an $ARCH(\infty)$ representation given by (5). Finally, notice that Assumption 1(b) allows for a non-zero unconditional correlation (endogeneity) between ε_t and e_t when Σ_v is non-diagonal. $\diamondsuit$

Remark 2.3. Under Assumption 1(b), (u_t) satisfies a functional central limit theorem (FCLT), such that the weak convergence result

$$B_n(r) := \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nr \rfloor} u_j \Rightarrow B(r) \quad \text{on } D[0,1] \tag{6}$$

holds, where $B(r)$ is a Brownian motion with variance $\omega^2 := C(1)^2 \sigma_e^2$, where $\sigma_e^2 := \mathbb{E}_{\mathcal{F}_{j-1}}\left(e_j^2\right)$. When $c \in \mathbb{R}$ in Assumption 1(a), denote by

$$J_c(t) := \int_0^t e^{c(t-s)} dB(s) \tag{7}$$

the Ornstein-Uhlenbeck [OU] process associated with the Brownian motion B in (6). $\diamondsuit$

As we shall see in Section 3, the IVX approach of KMS is based on a filtration of x_t, of the following form:

Definition 1. *Denote by $\tilde{z}_t$ a generalisation of the IVX instrument process of Phillips and Magdalinos (2009) generated from filtering the observed series x_t as follows*

$$\tilde{z}_t = \varphi_n \tilde{z}_{t-1} + \Delta x_t, \quad t \geq 1 \tag{8}$$

where $\Delta x_t := x_t - x_{t-1}$, with the recursion in (8) initialised at $\tilde{z}_0 := \tilde{z}_0(n)$, and where φ_n, the user-chosen autoregressive parameter in the recursion, satisfies the condition that $n(\varphi_n - 1) \to -\infty$.

Remark 2.4. The IVX instrument suggested by both Phillips and Magdalinos (2009) and KMS is a special case of Definition 1 where the instrument is initialised at $\tilde{z}_0 = 0$. KMS suggest setting the recursion parameter in (8) to $\varphi_n = 1 + aT^{-\gamma}$, for some $a < 0$ and $\gamma \in (0,1)$. The IVX scale and exponent parameters, a and γ, respectively, are tuning parameters set by the practitioner; KMS recommend setting $a = -1$ and $\gamma = 0.95$. $\diamondsuit$

In order to provide formal classes of permissible initial conditions, $X_0(n)$, and initialisations of the IVX instrument in Definition 1, $\tilde{z}_0(n)$, we first define the bivariate partial sum process,

$$[U_n(r), \zeta_n(r)] := \left(1 - \varphi_n^2\right)^{1/2} \left[\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} z_{t-1}\varepsilon_t, \sum_{t=1}^{\lfloor nr \rfloor} \varphi_n^{t-1}\varepsilon_t\right] \quad r \in [0,1] \tag{9}$$

where $z_t := \sum_{j=1}^{t} \varphi_n^{t-j} u_j$ denotes the (unobservable) series which obtains from a recursion of the form given in (8), but with Δx_t replaced by u_t, and initialised at zero. In Lemma 1 we next establish the joint convergence in distribution and the asymptotic independence of the random elements in (9) and the partial sum process in (6).

Lemma 1. *Under Assumption 1,* $[U_n(r), B_n(r), \zeta_n(r)] \Rightarrow [U(r), B(r), \zeta]$ *on* $D[0,1]^3$*, where* $U(r)$ *and* $B(r)$ *are independent Brownian motions with* $\mathbb{E}U(r)^2 = \omega^2\sigma_\varepsilon^2 r$*, and* $\zeta \stackrel{d}{=} N(0, \sigma_\varepsilon^2)$ *is independent of* $[U(r), B(r)]$*.*

Recalling the definition of the local-to-unit root parameter, c, from Assumption 1(a), we next define the key parameter, κ_n, whose value is determined by the order of magnitude of the unconditional variance of (x_t) for the value of the autoregressive parameter, ρ_n in (3), as follows:

$$\kappa_n := \left(1-\rho_n^2\right)^{-1} \mathbf{1}\{c = -\infty\} + n\mathbf{1}\{c \in \mathbb{R}\}. \tag{10}$$

Remark 2.5. Notice therefore that: $\kappa_n = n$ when (x_t) is a local-to-unit root process; $\kappa_n = O(n^\alpha)$ when (x_t) is a mildly integrated process (i.e. such that $\rho_n = 1 + d/n^\alpha$, $\alpha \in (0,1)$, with d a negative constant); and, $\kappa_n = O(1)$ in the weakly persistent case. $\diamondsuit$

Assumption 2. *Consider the random elements* $B_n(\cdot)$*,* $U_n(\cdot)$ *and* $\zeta_n(\cdot)$ *in (6) and (9).*

(a) (i) When $c = -\infty$*,* $X_0(n) = o_p(n^{1/2})$*.*

(ii) When $c \in \mathbb{R}$*,* $n^{-1/2}X_0(n) \xrightarrow{d} \mathbb{X}_0$*, where* $\mathbb{X}_0$ *is Gaussian and independent of* $\{[U(r), \zeta] : r \in [0,1]\}$*; convergence in distribution of* $\left[U_n(r), B_n(r), \zeta_n(r), n^{-1/2}X_0(n)\right]$ *holds on* $D[0,1]^4$*.*

(b) (i) When $c = -\infty$*,* $\tilde{z}_0(n) = o_p(n^{1/2})$*.*

(ii) When $c \in \mathbb{R}$*,* $n^{-1/2}\tilde{z}_0(n) = G_c\left(B_n, n^{-1/2}X_0(n)\right) + o_p(1)$ *for some* $\mathbb{P}_{B,\mathbb{X}_0}$*-a.s. continuous function* $G_c : D[0,1] \times \mathbb{R} \to \mathbb{R}$ *where* $\mathbb{P}_{B,\mathbb{X}_0}$ *is the distribution of* $(B, \mathbb{X}_0)$*.*

(c) $n^{-1/2}\kappa_n^{-1/2}(1-\varphi_n)^{-1/2}|\tilde{z}_0(n) - X_0(n)| = o_p(1)$*.*

Remark 2.6. In view of Lemma 1, the requirements of part (ii) of Assumption 2(a), relating to the strongly persistent case where $c \in \mathbb{R}$, consist of joint convergence in distribution of $[U_n(r), B_n(r), \zeta_n(r)]$ and $n^{-1/2}X_0(n)$ and independence of X_0 and $\{[U(r), \zeta] : r \in [0,1]\}$. Both of these requirements hold trivially if the limit $\mathbb{X}_0$ is non-random. A random limit $\mathbb{X}_0$ that satisfies Assumption 2(a) will typically be $\mathcal{F}_0$-measurable. We may construct such a class of initial conditions from an innovation process $(u_t)_{t\leq 0}$ that satisfies Assumption 1(b). If $(h_{n,t})$ is a deterministic triangular array, an initial condition

$$X_0(n) = \sum_{t=0}^{\infty} h_{n,t}u_{-t} = \sum_{t=0}^{\infty} \tilde{h}_{n,t}e_{-t}, \quad \tilde{h}_{n,t} := \sum_{k=0}^{t} h_{n,k}c_{t-k} \tag{11}$$

will satisfy $n^{-1/2}X_0(n) \xrightarrow{d} \mathbb{X}_0 \overset{d}{=} N\left(0, \tilde{h}^2\sigma^2\right)$ provided that

$$\frac{1}{n}\sum_{t=0}^{\infty} \tilde{h}_{n,t}^2 \to \tilde{h}^2 \quad \text{and} \quad \frac{1}{n}\sup_{t\geq 0} \tilde{h}_{n,t}^2 \to 0. \tag{12}$$

By (S.1) in the Appendix, $[U_n(r), B_n(r), \zeta_n(r)]' = \sum_{t=1}^{\lfloor nr \rfloor} \xi_{n,t} + o_p(1)$ where $\xi_{n,t}$ is a $\mathcal{F}_t$ -martingale difference array; since $\mathbb{E}_{\mathcal{F}_{t-1}}\left(e_{-t}\xi_{n,t}'\right) = e_{-t}\mathbb{E}_{\mathcal{F}_{t-1}}\left(\xi_{n,t}'\right) = 0$, the FCLT employed in Lemma 1 extends to $\left(\sum_{t=1}^{\lfloor nr \rfloor} \xi_{n,t}, X_0(n)\right) \Rightarrow (\xi, \mathbb{X}_0)$ and that $\mathbb{X}_0$ is independent of ξ. Hence, when $X_0(n)$ is given by (11), Assumption 2(a) will hold as long as Assumption 1(b) and (12) are satisfied. A leading case where this holds is where $X_0(n) = \sum_{i=0}^{\lfloor \tau n \rfloor} \rho_n^i u_{-i}$, $\tau \geq 0$, for $c < 0$. Here it is seen that while the (X_t) process satisfies the recursion in (3) starting at time $t = -\lfloor \tau n \rfloor$, it is only observed for $t \geq 0$. When $\tau = 0$, then $X_0(n) = u_0$, so that the initial condition is assumed to be an $O_p(1)$ random variable. When $\tau > 0$, $X_0(n)$ is of $O_p(n^{1/2})$; in particular, here $n^{-1/2}X_0(n) \Rightarrow \bar{J}_c(\tau)$, where $\bar{J}_c(s)$ denotes the OU process generated by $\mathrm{d}\bar{J}_c(s) = c\bar{J}_c(s) + \mathrm{d}\bar{B}(s)$, where $\bar{B}(s)$ is a Brownian motion with variance ω^2 and which is independent of $B(s)$. As $\tau \to \infty$, $X_0(n)$ approaches a draw from the unconditional distribution of (X_t). $\diamondsuit$

Remark 2.7. In view of part (i) of Assumptions 2(a) and 2(b), Assumption 2(c) is automatically satisfied when the rate of persistence of the regressor is the same or higher than that of the IVX instrument: $\kappa_n^{-1/2}(1-\varphi_n)^{-1/2} = O(1)$. If $\varphi_n > \rho_n$, the IVX limit distribution theory may accommodate an initialisation $X_0(n) = o_p(n^{1/2})$ only if an instrument initialisation $\tilde{z}_0(n)$ is chosen that is not too distant from $X_0(n)$: the most restrictive scenario arises under stationarity ($\kappa_n = 1$), where Assumption 2(c) postulates that the distance between $X_0(n)$ and $\tilde{z}_0(n)$ may diverge at a rate slower than $n^{1/2}(1-\varphi_n)^{1/2}$. The regressor and instrument initialisations $X_0(n)$ and $\tilde{z}_0(n)$ may only be chosen independently if the rate of both is restricted to $O_p\left(\kappa_n^{1/2}\right)$. This corrects an error in the initialisation assumption of KMS under stationarity ($\alpha = 0$ in the first paragraph of page 1510 of KMS): Assumption 2(c) above should be added to the maintained rate $X_0(n) = o_p(n^{1/2})$, or the latter needs to be restricted to $X_0(n) = O_p(1)$. In particular, given they initialise the IVX instrument at $\tilde{z}_0(n) = 0$, the limiting results given in KMS for the weakly dependent case require that $X_0(n) = O_p(1)$, and not the claimed $o_p(n^{1/2})$ rate for $X_0(n)$ stated on page 1510 of KMS. $\diamondsuit$

Remark 2.8. Assumption 2(b) delimits the class of allowable initialisations of the IVX instrument, $\tilde{z}_t$, of Definition 1. In particular, it allows for the case where

$$\tilde{z}_0(n) = x_0 - \bar{x}_n \tag{13}$$

in which $\bar{x}_n := n^{-1}\sum_{j=1}^{n} x_j$, which initialises the instrument at an estimate of the initial condition of x_t. We will later base our modified version of the KMS test on this initialisation. When $c = -\infty$, $X_0(n) = o_p\left(n^{1/2}\right)$ and $\bar{x}_n = O_p\left(n^{-1/2}\kappa_n\right) = o_p\left(n^{1/2}\right)$ (since $\kappa_n/n \to 0$) so part (i) of Assumption 2(b) is satisfied. When $c \in \mathbb{R}$,

$$\begin{aligned}\frac{1}{\sqrt{n}}\bar{x}_n &= \int_0^1 \frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} dr + \frac{1}{\sqrt{n}} X_0(n) \frac{1-\rho_n^n}{n(1-\rho_n)} + o_p(1) \\ &= \int_0^1 \left(B_n(r) + ce^{cr}\int_0^r e^{-cs} B_n(s)\, ds\right) dr - \frac{1}{\sqrt{n}} X_0(n) \frac{1-e^c}{c} + o_p(1)\end{aligned}$$

so Assumption 2(b) is satisfied with

$$G_c(x,y) = \int_0^1 \left(x(r) + ce^{cr}\int_0^r e^{-cs} x(s)\, ds\right) dr + y(1-e^c)/c.$$

Assumption 2(c) is satisfied by (13) since $\tilde{z}_0(n) - X_0(n) = \bar{x}_n$ and

$$n^{-1/2}\kappa_n^{-1/2}(1-\varphi_n)^{-1/2}|\bar{x}_n| = O_p\left(n^{-1}\kappa_n^{1/2}(1-\varphi_n)^{-1/2}\right) = o_p(1)$$

for $c \in \mathbb{R} \cup \{-\infty\}$. $\diamondsuit$

3 IVX Predictability Tests

In the context of the predictive regression model in (1)-(3), KMS develop the IVX-based test for the null hypothesis that $\beta = \beta_n$ based on the instrumental variable t-statistic,

$$T_n(\varphi_n) = N_n(\beta_n^* - \beta_n) \tag{14}$$

where

$$\beta_n^* := \left(\sum_{t=1}^n \underline{y}_t \tilde{z}_{t-1}\right)\left(\sum_{t=1}^n \underline{x}_{t-1}\tilde{z}_{t-1}\right)^{-1} \tag{15}$$

$$N_n := \hat{\sigma}_\varepsilon^{-1}\left(\sum_{t=1}^n \tilde{z}_{t-1}^2 - n\bar{z}_{n-1}^2\left(1-\hat{\rho}_{\varepsilon u}^2\right)\right)^{-1/2}\sum_{t=1}^n \underline{x}_{t-1}\tilde{z}_{t-1}. \tag{16}$$

in which $\underline{y}_t := y_t - n^{-1}\sum_{j=1}^n y_j$, $\underline{x}_{t-1} := x_{t-1} - n^{-1}\sum_{j=1}^n x_{t-1}$, $\bar{z}_{n-1} = n^{-1}\sum_{j=1}^n z_{t-1}$ and where $\hat{\sigma}_\varepsilon^2 := n^{-1}\sum_{t=1}^n \hat{\varepsilon}_t^2$ is a consistent estimate of the short run variance of ε_t; as discussed in KMS, a natural choice for $\hat{\varepsilon}_t$ are the OLS residuals obtained from estimating (1), that is from a OLS regression of y_t on x_{t-1}. The term $\hat{\rho}_{\varepsilon u}$ is an estimate of the long run correlation between ε_t and u_t; a discussion on the choice of estimator of this quantity in practice is given on pages 1513 and 1524 of KMS. Crucially, in the context of the t-statistic $T_n(\varphi_n)$ in (14), KMS initialise the IVX instrument, $\tilde{z}_t$, at $\tilde{z}_0 = 0$.

Because our interest in this paper is on testing the null of no predictability from x_{t-1}, we will restrict attention to the null hypothesis of no predictability, $H_0 : \beta = 0$, in what follows, though the results we give apply equally to tests of the more general null hypothesis that $\beta = \beta_n$.

As discussed in Section 2, if x_t is weakly, mildly, or strongly persistent with an initial condition that is, at most, of $O_p(1)$, $o_p(n^{\lambda/2})$ with $\lambda \in (0,1)$, or $o_p(n^{1/2})$, respectively, then KMS show that in large samples $T_n(\varphi_n)$ converges to a standard normal distribution under $H_0 : \beta = 0$. In the empirically most relevant case of strongly or mildly persistent predictors, this, however, rules out initial conditions generated according to Assumption 2 which permit the initial condition to be of $O_p(\kappa_n^{1/2})$, recalling that $\kappa_n = n$ in the strongly persistent case and $\kappa_n = n^\lambda$ in the mildly persistent case. In Section 4 we will show that $T_n(\varphi_n)$ continues to admit a standard normal limiting null distribution when x_t is either strongly or mildly persistent with an $O_p(\kappa^{1/2})$ initial condition. However, numerical simulations presented in Section 5 will show that in this scenario tests for predictability based on comparing $T_n(\varphi_n)$ to standard normal asymptotic (or bootstrap) critical values turn out to be severely undersized in finite samples, with this undersize more pronounced the larger is the value of the initial condition. This behaviour is also exhibited by another popular test for predictability, namely the Bonferroni Q test of Campbell and Yogo (2006); see Astill *et al.* (2024).

We, therefore, propose a simple modification to the IVX-based test procedure of KMS which, for various parametrisations of a strongly persistent predictor, has controlled size regardless of the magnitude of the initial condition. This modification involves calculating the $T_n(\varphi_n)$ test statistic exactly as in (14)-(16) but where the instrument $\tilde{z}_t$ is initialised not at zero but at $\tilde{z}_0 = x_0 - \bar{x}_n$ where $\bar{x}_n := n^{-1}\sum_{t=1}^n x_t$; that is, we initialise the instrument at an estimate of the initial condition of (x_t).[2] Henceforth we will denote our modified IVX t-statistic where the instrument is initialised at $\tilde{z}_0 = x_0 - \bar{x}_n$ as $T_n^\dagger(\varphi_n)$.

Remark 3.1. The test statistics outlined above can be used to test against either one-sided alternatives, $H_1 : \beta > 0$ or $H_1 : \beta < 0$, or against two-sided alternatives, $H_1 : \beta \neq 0$. For the latter case, one can equivalently use upper-tailed tests based on $T_n(\varphi_n)^2$ and $T_n^\dagger(\varphi_n)^2$, as noted by, *inter alia*, KMS. $\diamondsuit$

Remark 3.2. Although we have focussed attention on the case where the predictive regression in (1) contains a single candidate lagged predictor, x_{t-1}, the results we provide in this paper extend straightforwardly to the case where the predictive regression contains

[2]Another possible approach to consider would be to initialise the instrument at x_0 rather than $x_0 - \bar{x}_n$ when computing the $T_n^\dagger(\varphi_n)$ statistic. The instrument in this case would not, however, be invariant to the mean of x_t, μ, and so we will not use this initialisation.

multiple putative lagged predictors; that is,

$$y_t = \mu_y + \sum_{k=1}^{K} \beta_k x_{k,t-1} + \varepsilon_t, \qquad t = 1, ..., n \tag{17}$$

where $x_{k,t}$ is a set of K series each of which satisfies both Assumptions 1 and 2, with associated autoregressive coefficients $\rho_{k,n}$, $k = 1, ..., K$, which, as in KMS, are such that all of the series belong to the same persistence class (be it strongly, mildly or weakly persistent). In particular, it can be shown, along similar lines to the proof of Theorem 1 below, that the IV Wald statistic (where each of the K instruments is initialised at zero) proposed in Equation (19) of KMS (page 1514), W_{IVX}, for testing the joint null hypothesis that none of the series is predictive for y_t, $H_0 : \beta_1 = \cdots = \beta_K = 0$ in (17), admits a χ^2_K limiting null distribution provided the initial condition for each of the K series is of $o_p(n^{1/2})$ in the strongly persistent case, of $o_p(n^{\lambda/2})$ with $\lambda \in (0, 1)$ in the mildly persistent case, and of $O_p(1)$ in the weakly persistent case. Moreover, the generalisation of this statistic, $W^{\dagger}_{IVX}$ say, which initialises the K instruments at $\tilde{z}_{k,0} = x_{k,0} - \bar{x}_k$, where $\bar{x}_k := n^{-1} \sum_{t=1}^{n} x_{k,t}$, will have a χ^2_K limiting null distribution under the much weaker conditions that the initial conditions of the $x_{k,t}$ are of $O_p(n^{1/2})$ in the strongly persistent case, and of $o_p(n^{1/2})$ in the mildly or weakly persistent cases. $\diamondsuit$

Remark 3.3. As we will subsequently see in section Section 5, our Monte Carlo simulation experiments reveal that the $T^{\dagger}_n(\varphi_n)$ test, implemented with the residual wild bootstrap algorithm of Demetrescu *et al.* (2023), is able to control size regardless of the degree of persistence of the predictor or the magnitude of the initial condition, and avoids the potentially catastrophic power losses that the bootstrap implementation of the standard KMS test, $T_n(\varphi_n)$ suffers when the magnitude of the initial condition is not close to zero. Conversely, however, the standard KMS test is more powerful than the test based on $T^{\dagger}_n(\varphi_n)$ when the magnitude of the initial condition is close to zero and the predictor is close to being a pure unit root process. In order to simultaneously exploit the superior power of the test based on $T^{\dagger}_n(\varphi_n)$ when the initial condition is large and that of the test based on $T_n(\varphi_n)$ when the initial condition is small, we also propose a *union-of-rejections* based testing strategy: (i) for tests against $H_1 : \beta > 0$, this is equivalent to a test that rejects for large positive values of the statistic $U_R := \max\{T_n^\dagger(\varphi_n), T_n(\varphi_n)\}$; (ii) for testing against $H_1 : \beta < 0$, the test rejects for large negative values of $U_L := \min\{T_n^\dagger(\varphi_n), T_n(\varphi_n)\}$; (iii) for testing against $H_1 : \beta \neq 0$, the test rejects for large values of $U_{2S} := \max\{T_n^\dagger(\varphi_n)^2, T_n(\varphi_n)^2\}$. Details on how this *union-of-rejections* based testing strategy can be implemented in a size-controlled manner are outlined in the supplementary appendix. $\diamondsuit$

4 Asymptotic Results

Denote by

$$x_{0,t} := \sum_{j=1}^{t} \rho_n^{t-j} u_j \quad \text{and} \quad \tilde{z}_{0,t} := \sum_{j=1}^{t} \varphi_n^{t-j} \Delta x_{0,j} \tag{18}$$

the restrictions of x_t with $X_0(n) = 0$ and of $\tilde{z}_t$ with $X_0(n) = \tilde{z}_0(n) = 0$; recall the definition of z_t below (9). It is easy to see that (3) yields an autoregressive process

$$x_t = \mu(1-\rho_n) + \rho_n x_{t-1} + u_t$$

and backward recursion gives the following decompositions:

$$x_t = \mu + X_0(n)\rho_n^t + x_{0,t} \tag{19}$$

and

$$\tilde{z}_t = \tilde{z}_{0,t} + \varphi_n^t \tilde{z}_0(n) - X_0(n)(1-\rho_n)\frac{\varphi_n^t - \rho_n^t}{\varphi_n - \rho_n}. \tag{20}$$

A proof of the results in (19) and (20) can be found in Magdalinos and Petrova (2025). Defining $\underline{x}_{t-k} := x_{t-k} - n^{-1}\sum_{t=1}^{n-k} x_t$, for $k = 0, 1$, and $\bar{x}_{0,n-1} = n^{-1}\sum_{t=1}^{n-1} x_{0,t}$, we may then state the following lemma:

Lemma 2. *Under Assumptions 1 and 2, the following approximations hold as $n \to \infty$,*

(i) $n^{-1/2}(1-\rho_n^2\varphi_n^2)^{1/2}\left(\sum_{t=1}^n \tilde{z}_{t-1}\varepsilon_t - \sum_{t=1}^n \tilde{z}_{0,t-1}\varepsilon_t\right) = \frac{\tilde{z}_0(n)}{\sqrt{n}}\zeta_n(1) + o_p(1)$

(ii) $n^{-1}\left(1-\rho_n^2\varphi_n^2\right)\left(\sum_{t=1}^n \underline{x}_t \tilde{z}_t - \sum_{t=1}^n x_{0,t}\tilde{z}_{0,t}\right) = -2\frac{\tilde{z}_0(n)}{\sqrt{n}}\frac{\bar{x}_{0,n-1}}{\sqrt{n}} + 2\frac{X_0(n)}{\sqrt{n}}\psi_n(c) + o_p(1)$ *where* $\psi_n(c) = 0$ *when* $c = -\infty$ *and*

$$\begin{aligned}\psi_n(c) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n-1}\rho_n^j u_j + \frac{1-e^c}{c}\frac{1}{\sqrt{n}}\sum_{t=1}^n u_t \\ &\quad -\frac{c}{n^{3/2}}\sum_{t=1}^n \rho_n^{t-1}x_{0,t-1} + (1-e^c)\frac{1}{n^{3/2}}\sum_{t=1}^n x_{0,t-1} + \frac{\tilde{z}_0(n)}{\sqrt{n}}\left(1+\frac{1-e^c}{c}\right) \\ &\quad +\frac{X_0(n)}{\sqrt{n}}(1-e^c)\left(\frac{1}{2}(1+e^c)+\frac{1}{c}(1-e^c)\right) \qquad (21)\end{aligned}$$

when $c \in \mathbb{R}$.

(iii) $n^{-1}\left(1-\rho_n^2\varphi_n^2\right)\left(\sum_{t=1}^n \tilde{z}_t^2 - \sum_{t=1}^n \tilde{z}_{0,t}^2\right) = \left(n^{-1/2}\tilde{z}_0(n)\right)^2 + o_p(1)$.

Remark 4.1. Noting that $\zeta_n \xrightarrow{d} \zeta$ and that $n^{-1/2}\bar{x}_{0,n-1} = n^{-3/2}\sum_{t=1}^{n-1} x_{0,t}$ converges in distribution to $\int_0^1 J_c(t)\,dt$ when $c \in \mathbb{R}$ and to 0 when $c = -\infty$, the approximations in parts (i)-(iii) of Lemma 2 will be $o_p(1)$ if $n^{-1/2}\tilde{z}_0(n) = o_p(1)$; by Assumption 2.(b), $n^{-1/2}\tilde{z}_0(n) = o_p(1)$ whenever x_t is not a local to unity process (i.e. when $c = -\infty$). Hence, $\tilde{z}_0(n)$ will have an asymptotically non-negligible effect on the IVX estimator only in the case where the regressor x_t is a local-to-unit root process. In the local-to-unit root case, while both the initial condition of the regressor and the instrument initialisation contribute to the limiting distribution of the IVX estimator when these are of the form given in Assumption 2, it turns out that they do so in such a way that maintains the asymptotically mixed Gaussian property for the IVX estimator (see Remark 4.2) and, consequently, the standard normal asymptotic null distribution of the IVX t-statistic. ♢

The preservation of the standard normal asymptotic null distribution of the IVX-based t-statistic, $T_n(\varphi_n)$, under initial conditions that are not asymptotically negligible, and for the corresponding t-statistic, $T_n^\dagger(\varphi_n)$, based on initialisations of the IVX instrument other than zero is the main result of this paper. We now formally state these large sample results in Theorem 1, along with the limiting null distribution of the statistic formed from the maximum of these two statistics; cf. Remark 3.3.

Theorem 1. *Let data be generated according to (1)-(3). If Assumptions 1 and 2 hold, then under $H_0 : \beta = 0$ we have the joint convergence result,*

$$\left[T_n(\varphi_n), T_n^{\dagger}(\varphi_n)\right] \xrightarrow{d} \left[T, T^{\dagger}\right]$$

where T and $T^{\dagger}$ are $N(0,1)$ random variables. Moreover, under weak persistence, $c = -\infty$, $T_{\max} := \max\left\{T, T^{\dagger}\right\} \stackrel{d}{=} N(0,1)$. Under strong persistence, $c \in \mathbb{R}$, $T_{\max}$ has density function

$$f_{T_{\max}}(\lambda) = \phi(\lambda)\,\mathbb{E}\left[|G_c(B, \mathbb{X}_0)| + |G_c(B, \mathbb{X}_0)|\,\Phi\left(\frac{\lambda}{|G_c(B, \mathbb{X}_0)|}\left(\left(\omega^2 + G_c(B, \mathbb{X}_0)^2\right)^{1/2} - \omega\right)\right)\right]$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the $N(0,1)$ density and distribution functions, respectively, B is Brownian motion with variance ω^2 and $G_c(B, \mathbb{X}_0)$ is the continuously distributed random variable defined in Assumption 2.

Corollary 1. *Let the conditions of Theorem 1 hold. Then under under $H_0 : \beta = 0$, the joint convergence result*

$$\left[T_n(\varphi_n)^2, T_n^{\dagger}(\varphi_n)^2\right] \xrightarrow{d} \left[T^2, T^{\dagger 2}\right]$$

holds, where T^2 and $T^{\dagger 2}$ are $\chi^2(1)$ random variables.

Remark 4.2. As discussed in Remark 4.1, the property that underlies Theorem 1 is the asymptotic mixed Gaussianity of the IVX estimator, so it is worth providing an insight as to why this property is not distorted by the contributions of the autoregressive and instrument initialisations in the local-to-unit root case. Firstly, note that both $n^{-1/2}\tilde{z}_0(n)$ and $\psi_n(c)$ in (21) are deterministic functionals of $\left(B_n, n^{-1/2}X_0(n)\right)$; hence part (ii) of Lemma 2 implies that the denominator of the IVX estimator, $n^{-1}\left(1 - \varphi_n^2\right)\sum_{t=1}^n \underline{x}_t \tilde{z}_t$, is asymptotically equivalent to a deterministic functional, $\Psi_c\left(B_n, n^{-1/2}X_0(n)\right)$ say, of $\left(B_n, n^{-1/2}X_0(n)\right)$. By parts (i) and (ii) of Lemma 2, when $c \in \mathbb{R}$, the IVX estimator can be written as

$$n^{1/2}\left(1 - \varphi_n^2\right)^{-1/2}\left(\tilde{\beta}_n - \beta\right) = \frac{1}{\Psi_c\left(B_n, n^{-1/2}X_0(n)\right)}\left[1, n^{-1/2}\tilde{z}_0(n)\right]\begin{bmatrix} U_n(1) \\ \zeta_n(1) \end{bmatrix} + o_p(1). \tag{22}$$

By virtue of Lemma 1 and Assumption 2, $[U_n(1), \zeta_n(1)]$ is asymptotically independent of $\left(B_n, n^{-1/2}X_0(n)\right)$ (and, hence, also of $n^{-1/2}\tilde{z}_0(n)$) and converges in distribution to a

Gaussian random vector. Consequently, the right side of (21) converges in distribution to a zero mean mixed Gaussian random variable. $\diamondsuit$

Remark 4.3. An expression for the limiting null distribution of the U_R union-of-rejections statistic defined in Remark 3.3 can be obtained using the formulae for the density function of $T_{\max} = \max\left\{T, T^{\dagger}\right\}$ (in the weak and strong persistence case) in Theorem 1. These formulae can also be used to obtain an expression for the limiting null distribution of the U_L statistic, noting that $\min(a,b) = -\max(-a,-b)$. $\diamondsuit$

5 Numerical Simulations

In this section we report the results from a Monte Carlo simulation exercise examining the empirical performance of the tests discussed in this paper.

Data are generated according to (1)-(3), setting $\mu_y = \mu = 0$, without loss of generality, for a sample of size $n = 250$; corresponding results for $n = 1000$ are qualitatively similar and can be found in the supplementary appendix. The autoregressive coefficient was set as $\rho_n = 1 + c/n$ for $c \in \{0, 2, 5, 10, 30, 40, 50\}$ (such that, for $n = 250$, ρ_n takes values between 1.0 and 0.8), while the innovations are generated from an IID sequence of bivariate standard normal variates; that is,

$$\begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix} \sim IID\ N(\mathbf{0}, \mathbf{\Sigma}), \quad \mathbf{\Sigma} = \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix} \tag{23}$$

with the (short run) endogeneity correlation in (23) set to $\delta = -0.95$.[3]

[3]The value of $\delta = -0.95$ is chosen because for many empirical datasets a large negative correlation is found between the innovations to returns and the innovations to the predictor. Indeed, in the two empirical examples we later report in Section 6 that use the earnings-price ratio and dividend-price ratio as predictors the full sample estimates of δ found by Campbell and Yogo (2006) are given by -0.983 and -0.948, respectively.

We generate the initial condition of (x_t) as[4]

$$X_0(n) = \alpha\sqrt{\frac{1}{1-\rho_n^2}}\mathbb{I}(c<0) \tag{24}$$

with the magnitude of the initial condition, α, varied among $\alpha \in \{0, 1, 3\}$. When $c = 0$, (24) sets $X_0(n) = 0$; this is without loss of generality, because all of the tests are exact invariant to $X_0(n)$ when $c = 0$. When $\alpha = 0$ and $c > 0$ the initialisation satisfies the conditions of, *inter alia*, KMS which require the initial condition to be at most of $o_p(n^{1/2})$. However, when $\alpha > 0$ and $c > 0$ the initial condition is proportional to the standard deviation of the X_t and is therefore of $O_p(n^{1/2})$. In all cases we test the null of no predictability, and examine both the size of the tests when $\beta = 0$ and under the local alternative $\beta = b/n$, $b \in \mathbb{R}$.

We report results for both one-sided $T_n(\varphi_n)$ and $T_n^\dagger(\varphi_n)$ and two-sided implementations of the $T_n(\varphi_n)^2$ and $T_n^\dagger(\varphi_n)^2$ tests, together with the corresponding union-of-rejections tests outlined in Remark 3.3, using critical values obtained using the residual wild bootstrap procedure of Demetrescu *et al.* (2023), outlined in Section S.2 of the supplementary appendix, with the autoregressive lag order for x_t determined using the Bayes Information Criterion with maximum lag order $p_{\max} = \lfloor 4(T/100)^{0.25}\rfloor + 1$. When constructing the IVX instrument in (8) we follow KMS and set $\varphi_n = 1 - 1/n^\gamma$ and consider values of $\gamma \in \{0.95, 0.75, 0.50\}$ for the $T_n(\varphi_n)$ and $T_n(\varphi_n)^2$ test statistics and $\gamma \in \{0.95, 0.75\}$ for the $T_n^\dagger(\varphi_n)$ and $T_n^\dagger(\varphi_n)^2$ test statistics. Henceforth, we will refer to tests based on $T_n(\varphi_n)$ and $T_n(\varphi_n)^2$ where the IVX instrument is constructed using an exponent γ generically as IVX_γ and tests based on $T_n^\dagger(\varphi_n)$ and $T_n^\dagger(\varphi_n)^2$ where the IVX instrument is constructed using an exponent γ as $IVX_\gamma^\dagger$.

[4] Results are only reported for $\alpha = 0$ and positive values of α, as the simulation results for $\alpha = -1, -3$ were essentially identical to those for $\alpha = 1, 3$, respectively. We also repeated the simulations for random initial conditions generated according to $X_0(n) = Z \times \alpha\sqrt{\frac{1}{1-\rho_n^2}}\mathbb{I}(c<0)$, where $Z \sim N(0,1)$ for $\alpha \in \{1, 3\}$. For a given value of α, these results were qualitatively similar, although not identical, to those for reported for a fixed initial condition with the same α. We also performed simulations for the case where the DGP was initialised at time $t = -\lfloor \tau n\rfloor$, but only observed from $t = 0$; cf. Remark 2.6. As expected, as the value of τ was increased these results quickly began to closely mirror those for the random initial condition defined above with $\alpha = 1$, where $X_0(n)$ is a draw from the unconditional distribution of (X_t).

For the union-of-rejections test we always utilise an IVX exponent of $\gamma = 0.95$ for the constituent $T_n(\varphi_n)$ test statistic as recommended by KMS. We then consider two versions of the union-of-rejections test, one where the $T_n^{\dagger}(\varphi_n)$ test statistic is computed using $\gamma = 0.95$, denoted $U_{0.95}$, and one where the $T_n^{\dagger}(\varphi_n)$ test statistic is computed using $\gamma = 0.75$, denoted $U_{0.75}$.

To compute $\hat{\rho}_{\varepsilon u}$, the estimated long run correlation between ε_t and u_t in (16), we follow KMS and use long run variance and covariance estimators calculated using the Bartlett kernel with truncation lag $\lfloor n^{1/3} \rfloor$.

All simulations are performed using the RNDN function of Gauss 22.2 for 5,000 Monte Carlo replications, with the bootstrap critical values computed using $B = 499$ bootstrap replications.

We first examine the performance of the tests when $c = 0$ for both right tailed and left tailed alternatives using both one-sided and two sided tests. These results are reported in Figure 2. We see that for both right-tailed and left-tailed alternatives, and for one-sided and two-sided testing, that the best overall performance is given by the $IVX_{0.95}^{\dagger}$ and $U_{0.95}$ tests, with the $IVX_{0.75}^{\dagger}$ and $U_{0.75}$ tests close behind. For the tests proposed in KMS we see that power is increasing in the value of the exponent used to construct the IVX instrument, with the $IVX_{0.95}$ test being the best performing of the tests proposed by KMS.

We next look at the performance of the tests when testing against right tailed alternatives using both one-sided and two sided tests. Results are reported for $c \in \{2, 20, 50\}$ in Figures 3-5, with additional results for $c \in \{5, 10, 30, 40\}$ reported in Section S.3 of the supplementary appendix. We concentrate first on the one sided tests when $\alpha = 0$. In this scenario the best overall performance is displayed by the $IVX_{0.95}$ test, with this test showing controlled size and the best overall power profile across the values of c considered. Power for the $IVX_{0.75}$ test is lower than that of the $IVX_{0.95}$ test, with the power of the $IVX_{0.50}$ test lower still. This power ordering among the IVX_{γ} tests echoes the simulation results in KMS. Turning to the $IVX_{\gamma}^{\dagger}$ tests, we see that in general these display lower power than the corresponding IVX_{γ} test, though for $c = 2$ their power functions cross in the region of $b = 7$ both for $\gamma = 0.95$ and $\gamma = 0.75$. The difference between the powers of the IVX_γ and $IVX_\gamma^\dagger$, for a given value of γ, in general reduces as c increases, with essentially no difference between the tests' power functions when $c = 50$.

Turning to the case where $c > 0$ and $\alpha > 0$ a very different pattern emerges. We see that the IVX_γ tests become, often quite severely, undersized, with this undersize increasing in the values of c and α. We also see that this undersize is more severe the larger is the value of γ. As a consequence, the power of the IVX_γ test is severely reduced relative to the case where $\alpha = 0$. The $IVX_\gamma^\dagger$ tests, on the other hand, do not suffer from any noticeable undersize, and consequently are much more powerful than the IVX_γ tests, with $IVX_{0.95}^\dagger$ in particular displaying excellent power properties.

The contrasting power orderings between the IVX_γ and $IVX_\gamma^\dagger$ tests for $\alpha = 0$ *vis-à-vis* $\alpha > 0$ suggest that a union-of-rejections type testing strategy should be useful, allowing us to exploit the superior power of the $IVX_{0.95}$ test where the magnitude of the initial condition is small, and the superior power of the $IVX_{0.95}^\dagger$ test where the magnitude is large. The results in Figures 3–5 confirm this with $U_{0.95}$, the right-tailed union-of-rejections test based on $IVX_{0.95}$ and $IVX_{0.95}^\dagger$, being among the best performing tests for all values of c and α. We therefore recommend the use of this test in practice when uncertainty exists over the magnitude of the initial condition when testing in the right tail when $\delta < 0$. Results for two-sided variants of the tests against right-tailed alternatives are also reported in Figures 3-5 and the relative performance of the tests is qualitatively similar to that for the one-sided variants, with again the two-sided $U_{0.95}$ test recommended.

We turn next to the performance of the tests when testing against left-tailed alternatives using both one-sided and two-sided tests. These are reported in Figures 6-7 for $c \in \{2, 20\}$, with results for the additional values of c again reported in the supplementary appendix. We again begin by examining the performance of the one-sided tests when $\alpha = 0$. These results are slightly different to those for the right-tailed testing scenario discussed above, as now the best overall power performance is displayed by the U_γ tests. For $\alpha > 0$, the IVX_γ tests, again, become undersized and lacking in power, with this impact more pronounced the larger is the value of γ or α. For $\alpha > 0$ the best power performance is again displayed by the U_γ tests, although for $c = 2$ and $\alpha = 3$ the $U_{0.95}$ test displays significant oversize, and so overall we recommend the use of the $U_{0.75}$ test when testing in the left tail when $\delta < 0$. For two-sided testing, the $U_{0.95}$ test has one of the strongest overall power profiles for left-tail alternatives, and given that this test also has one of the strongest power profiles against right-tail alternatives we recommend the use of the $U_{0.95}$ test for two-tailed testing.

Overall, when $\delta < 0$ and there is uncertainty over the magnitude of the initial condition of the predictor, we recommend the use of $U_{0.95}$ for right-tailed one-sided tests and two-tailed tests, and $U_{0.75}$ test for one-sided left-tailed tests.

All of our simulation experiments discussed above pertain to the case where δ is negative. For predictive regressions generated according to (1)-(3), it is noted by, *inter alia*, Campbell and Yogo (2006, p.30) that replacing x_t with $-x_t$ in (1) flips the sign of both β and δ, such that a right-tailed test with $\delta < 0$ is asymptotically equivalent to a left-tailed test with $\delta > 0$. Therefore for $\delta > 0$ we recommend using $U_{0.95}$ for left-tailed one-sided tests and two-tailed tests, and $U_{0.75}$ test for one-sided right-tailed tests. Such a strategy is feasible in practice, given that δ can be consistently estimated by, for example, the OLS estimate given in Harvey *et al.* (2021, p.205). The possibility that the estimate of δ will have the wrong sign should only occur in cases where δ is close to zero, and here unreported simulation results show that there is little difference between the power of the $U_{0.95}$ and $U_{0.75}$ tests.

6 Empirical Illustration

In this section we apply the tests discussed in this paper to the dataset originally used by Campbell and Yogo (2006) to illustrate how the magnitude of the initial condition of the predictor can play a key role in determining whether a test signals a rejection. Following the motivating example in Section 1, we examine the predictability of monthly CRSP returns for sample data covering the period 1926M12–1994M12 when using either the dividend-price or earnings-price ratio as a predictor. Full data descriptions are provided in Campbell and Yogo (2006). The data were obtained from `https://sites.google.`

com/site/motohiroyogo/research/asset-pricing. We focus on these two predictors as they are arguably the most widely examined predictors for returns in the literature. The estimate of the endogeneity correlation, δ, for the dividend-price ratio is equal to -0.948 and for the earnings-price ratio it is -0.983, with these values close to the value of $\delta = -0.950$ used in our Monte Carlo simulation experiments in Section 5.

For each predictor we apply right-sided bootstrap implementations of the $IVX_{0.95}$, $IVX^{\dagger}_{0.95}$ and $U_{0.95}$ tests, performed using the exact same settings as in the Monte Carlo simulation exercise in Section 5, except that $B = 1,999$ bootstrap replications were used. We calculate bootstrap p-values for the $IVX_{0.95}$, $IVX^{\dagger}_{0.95}$ and $U_{0.95}$ tests, sequentially increasing the start date of the sample data used by one period. In particular, we first calculate the test statistics and associated bootstrap p-values over the full sample of data available, 1926M12–1994M12 so that 1926M12 is the date of initial value of the predictor, then for data over the sample period 1927M1–1994M12 where the initial value of the predictor is 1927M1, and so on, finishing at the sample period 1945M12–1994M12. For all of the tests, the bootstrap p-values are calculated as outlined in Section S.2 of the supplementary appendix.

For each start date we also compute an estimate of the magnitude of the initial condition proposed by Harvey and Leybourne (2005) given by

$$|\hat{\alpha}| := |x_0 - \hat{\mu}|/\hat{\sigma}_w \tag{25}$$

where $\hat{\mu} := T^{-1}\sum_{t=1}^{T} x_t$ and $\hat{\sigma}^2_w := T^{-1}\sum_{t=1}^{T}(x_t - \hat{\mu})^2$. Harvey and Leybourne (2005) show that while $|\hat{\alpha}|$ is not a consistent estimate of α, a monotonic relationship holds between $|\hat{\alpha}|$ and the true magnitude of the initial condition. The estimates of $|\hat{\alpha}|$ give an indication of how the relative magnitude of the initial condition is changing across start dates.

Figure 8 reports these results for the case were the lagged dividend-price ratio is used as a candidate predictor. In each subfigure the red dots indicate p-values that are greater than or equal to 0.05, signalling non-rejection of the null of no predictability at the 5% significance level, and green dots indicate p-values that are less than 0.05, signalling rejection of the null of no predictability at the 5% level. Additionally, the grey shaded areas highlight start dates for which a test fails to reject the null of no predictability. We see that the $IVX_{0.95}$ test rejects the null at the 5% level for 55.0% of the start dates considered, but with a long string of non-rejections between 1931-1933 and 1941-1942, as well as some dates in 1937-1938, that coincide with large estimated values of the initial condition of the predictor. This is in accord with the simulation results reported in Section 5 where we saw that this test is undersized and lacking in power when the initial condition of the predictor is large. The $IVX^{\dagger}_{0.95}$ test rejects the null less often than the $IVX^{\dagger}_{0.95}$ test for only 44.5% of start dates, but importantly this test is seen to reject for the start dates 1932M4-1932M6, 1937M12-1938M5 and 1941M2-1942M10, all of which are associated with large estimated initial conditions and, with the exception of 1938M2, are all start dates for which the $IVX_{0.95}$ test does not reject. Our preferred $U_{0.95}$ test rejects the null for 67.7% of start dates, which is a significantly higher proportion of start dates than either of the constituent tests used in its construction. Indeed, with the exception of 1937M11, the $U_{0.95}$ test rejects for each and every start date where either the $IVX_{0.95}$ or $IVX^{\dagger}_{0.95}$ tests reject, and for this lone start date $p(U_{0.95}) = 0.055$ so the non-rejection is marginal. These results clearly highlight the advantage of the union-of-rejections approach as the $U_{0.95}$ test is able to reject much more consistently across various initial condition magnitudes than either $IVX_{0.95}$ or $IVX^{\dagger}_{0.95}$ are able to do in isolation.

Figure 9 reports results when using the lagged earnings-price ratio as a predictor. For this predictor the $IVX_{0.95}$ test rejects the null of no predictability at the 5% significance level for 72.1% of the start dates considered, with clusters of non-rejections for start dates between 1931-1935 and 1941-1943 for which the estimated magnitude of the initial condition is again seen to be large. The $IVX^{\dagger}_{0.95}$ test, on the other hand, rejects for 85.6% of the start dates considered, including the periods between 1931-1935 and 1941-1943 where the $IVX_{0.95}$ test fails to reject. As with the results for the dividend-price ratio, there are again instances where the estimated initial condition is small and the $IVX^{\dagger}_{0.95}$ test fails to reject while the $IVX_{0.95}$ tests rejects. The $U_{0.95}$ union-of-rejections test again has the highest rejection frequency of the tests, rejecting the null of no predictability at the 5% significance level for 90.8% of the start dates considered.

Finally, Figure 10 plots, for each predictor, the estimated magnitude of the initial condition across start dates as well as the p-value of all three tests. This figure re-enforces the fact, for the earnings-price ratio in particular, that the p-value of the $U_{0.95}$ test is far less sensitive to the magnitude of the initial condition than the other two tests which explains the greater rejection frequency found for this test. Overall, the findings of this empirical study accord with the findings from our simulation exercise in Section 5 and give further weight to our recommendation that practitioners use the $U_{0.95}$ test in practice.

7 Conclusions

We have shown that the IVX predictability test of KMS can be very sensitive in practice to the magnitude of the initial condition (defined as the deviation of the starting value of a process from its unconditional mean) of the candidate predictor(s) under test. In particular, it can suffer from very substantial finite sample power losses for strongly persistent predictors which have an initial condition whose magnitude is large. To address this issue we have proposed a simple modification to the IVX test in which the instrument used to construct the IVX test is initialised not as in KMS at zero, but at the difference between the initial value and sample mean of the predictor. We have shown that this modified test retains the same limiting null distribution as the original IVX statistic but allows for a much wider class of initial conditions than is allowed for validity of the original KMS test. However, the modified test can have power below that of the original IVX test when the initial condition is small. We therefore proposed a union-of-rejections test, formed from these two tests, that was demonstrated to show the best overall power profile across a range of simulation DGPs. Finally, an empirical application to data originally used by Campbell and Yogo (2006) demonstrated the value of our preferred union-of-rejections based approach.

References

Arvanitis, S. and Magdalinos, T. (2018). Mildly explosive autoregression under stationary conditional heteroskedasticity. Journal of Time Series Analysis 39, 892-908.

Astill, S., Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R. (2024). Bonferroni Type Tests for Return Predictability and the Initial Condition. Journal of Business and Economic Statistics 42, 499-515.

Campbell, J. Y (1987). Stock returns and the term structure. Journal of Financial Economics 18, 373-399.

Campbell, J. Y. and Shiller, R.J. (1988a). Stock prices, earnings, and expected dividends. Journal of Finance 43, 661-676.

Campbell, J. Y. and Shiller, R.J. (1988b). The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195-228.

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.

Demetrescu, M., Georgiev I., Rodrigues P.M.M. and Taylor, A.M.R. (2023). Extensions to IVX methods of inference for return predictability, Journal of Econometrics, 237, Issue 2, Part C.

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

Fama, E.F. (1990). Stock returns, expected returns, and real activity. Journal of Finance 45, 1089-1108.

Fama, E.F. and French, K.R. (1988). Dividend yields and expected stock returns. Journal of Financial Economics 22, 3-24.

Fama, E.F. and French, K.R. (1989). Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25, 23-49.

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. (2021). Simple tests for stock return predictability with good size and power properties, Journal of Econometrics 224, 198–214.

Keim, D.B. and Stambaugh, R.F. (1986). Predicting returns in the stock and bond markets. Journal of Financial Economics 17, 357-390.

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

Magdalinos, T. and Petrova, K. (2025). Uniform and distribution-free inference with general autoregressive processes. Federal Reserve Bank of New York Staff Reports, no. 1151, April 2025. https://doi.org/10.59576/sr.1151

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

Nelson, C.R. and Kim, M.J. (1993). Predictable stock returns: the role of small sample bias. Journal of Finance 48, 641-661.

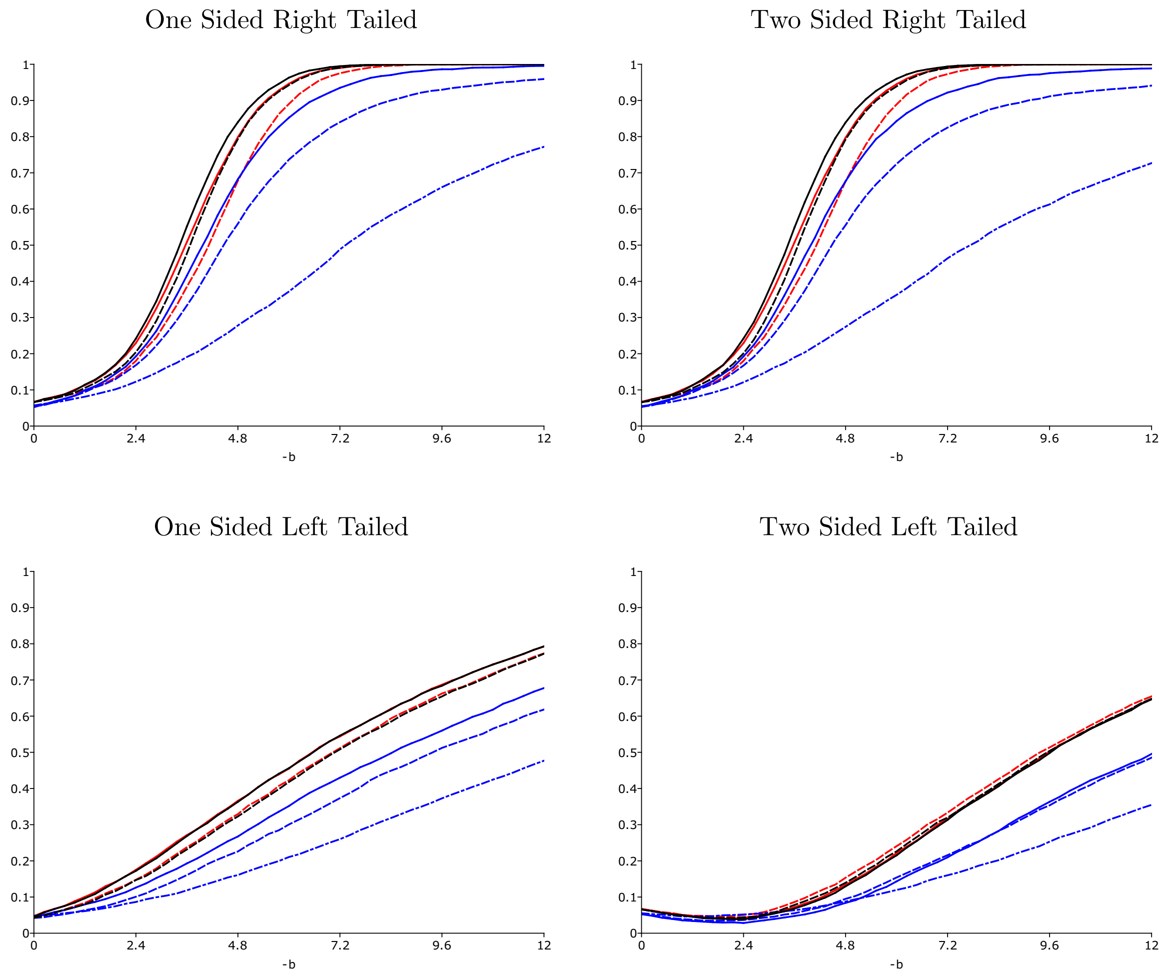
Phillips, P.C.B. and Magdalinos, T. (2009). Unit Root and Cointegrating Limit Theory When Initialization Is in the Infinite Past. Econometric Theory 25, 1682-1715.

Smeekes, S. and Taylor, A.M.R. (2012). Bootstrap union tests for unit roots in the presence of nonstationary volatility. Econometric Theory 28, 422-456.

Stambaugh, R.F. (1999). Predictive regressions. Journal of Finance 54, 375-421.

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

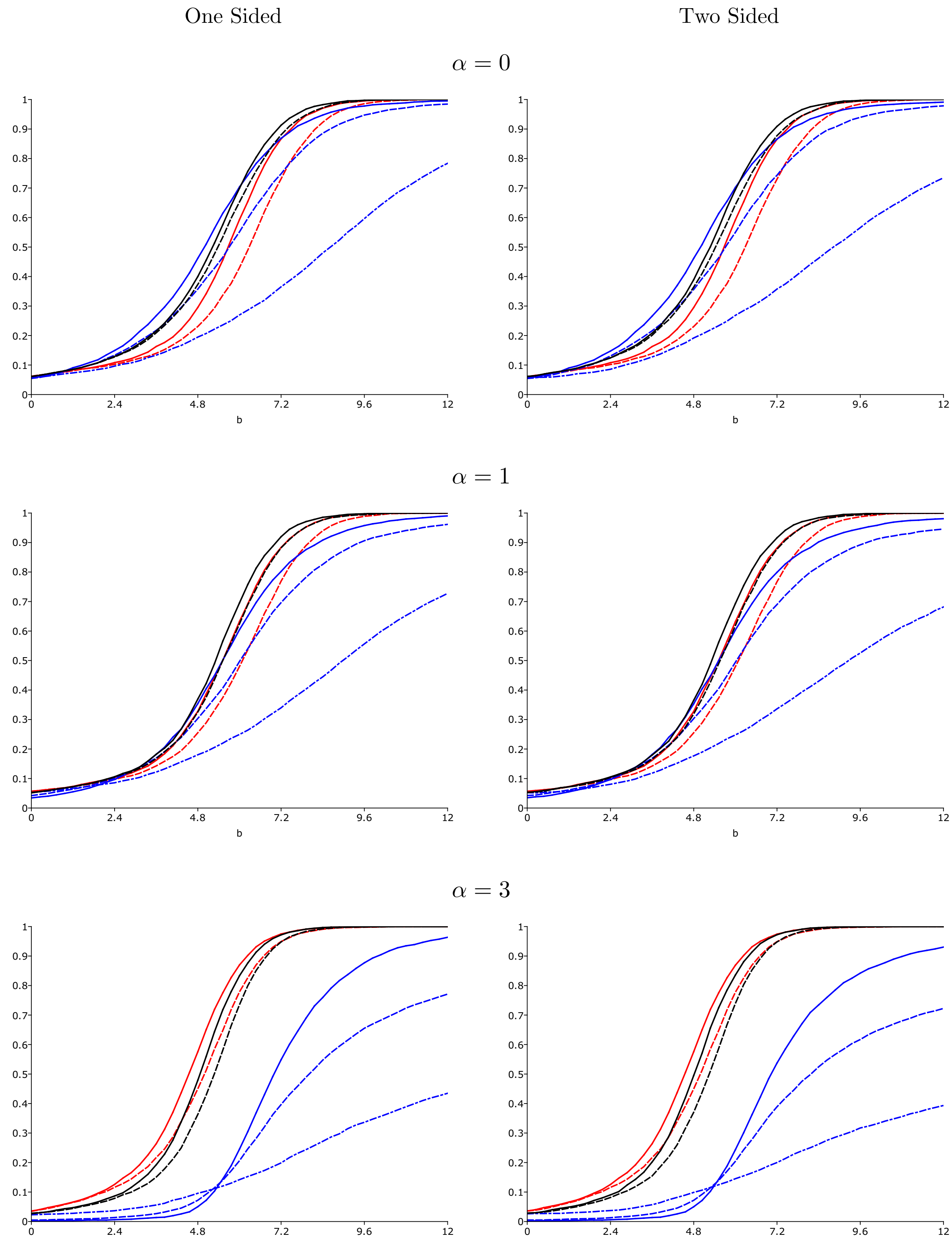
Figure 2: Finite Sample Power, $n = 250$, $\delta = -0.95$, $c = 0$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

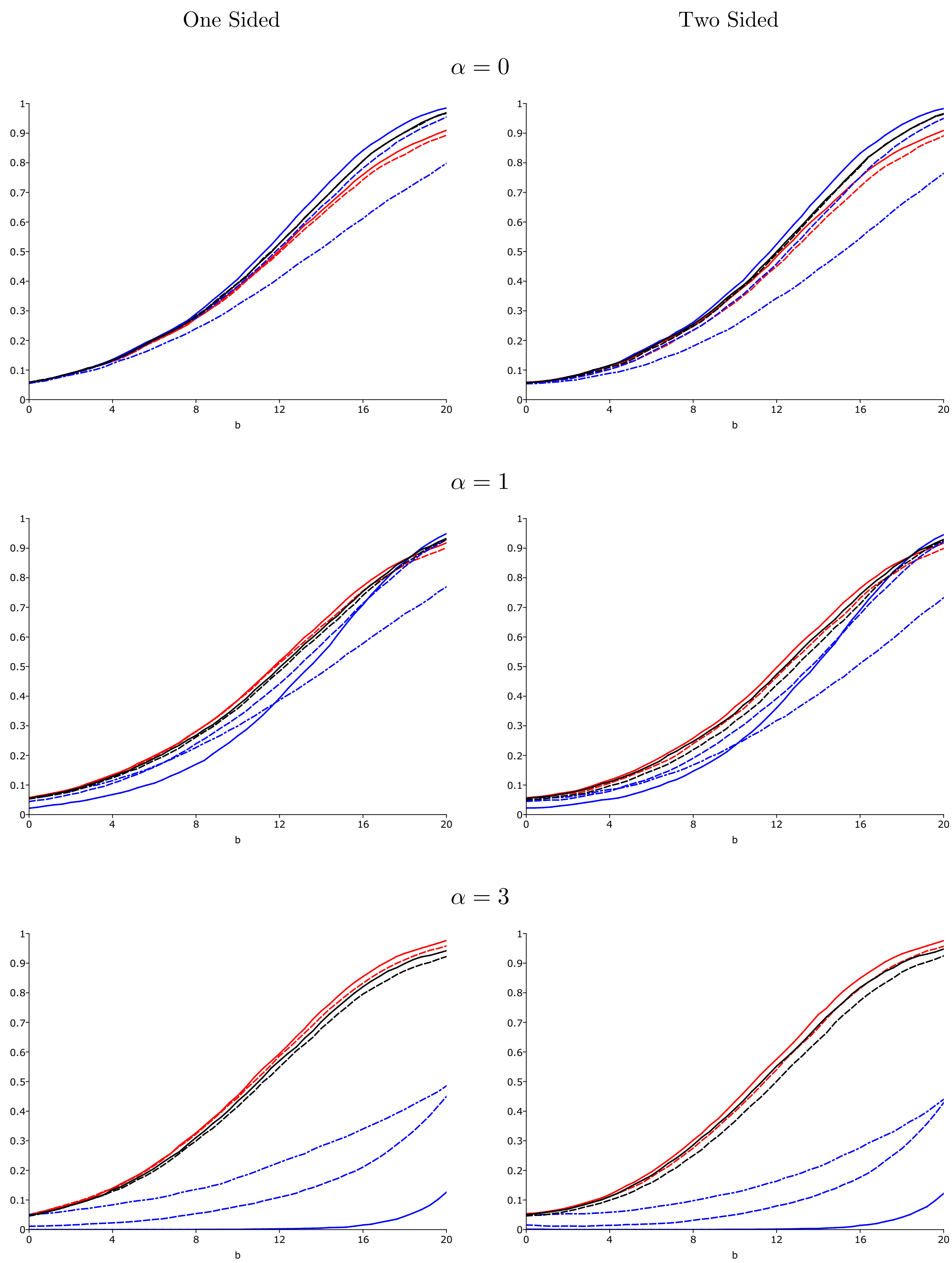
Figure 3: Finite Sample Power, Right Tail, $n = 250$, $\delta = -0.95$, $c = 2$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

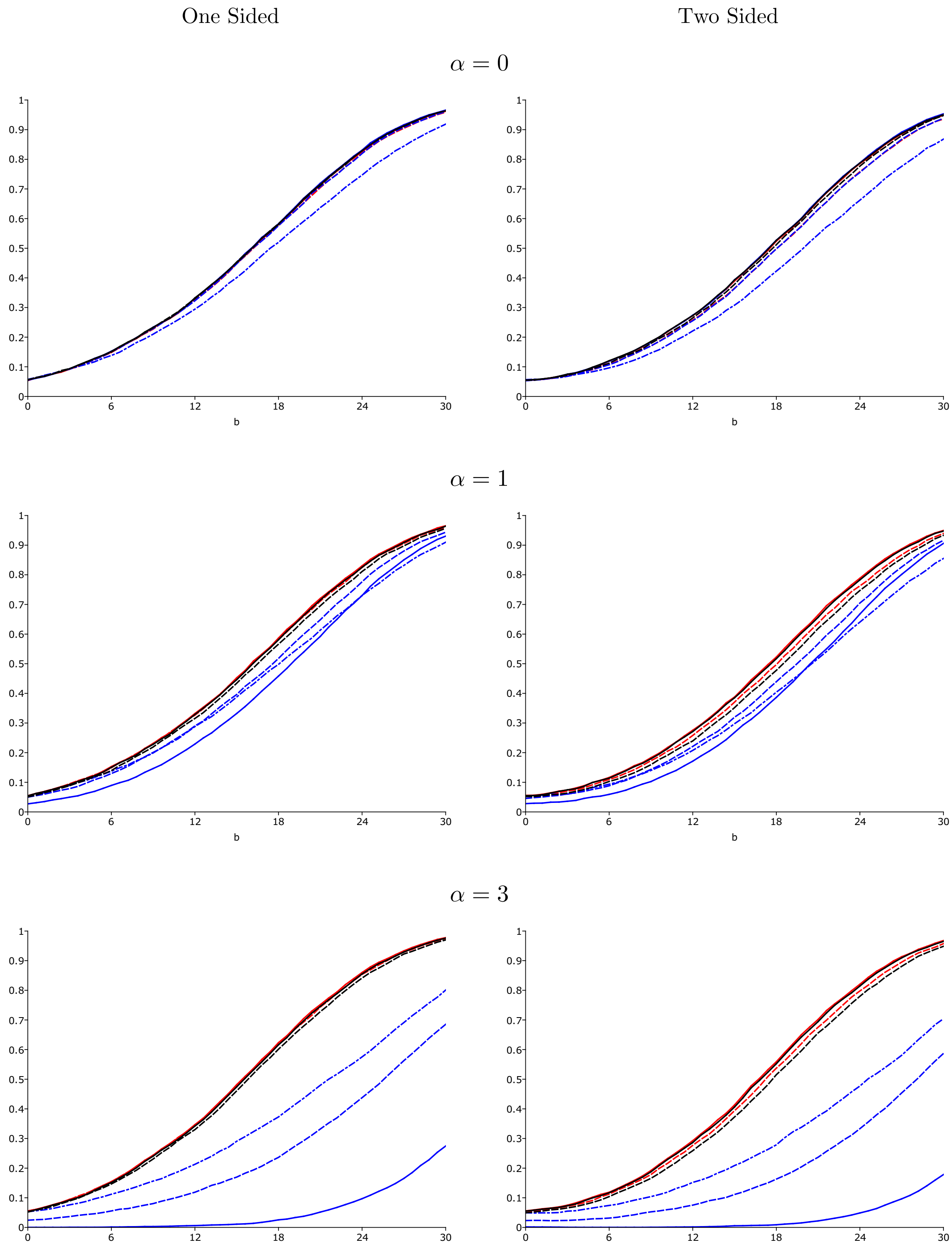
Figure 4: Finite Sample Power, Right Tail, $n = 250$, $\delta = -0.95$, $c = 20$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

Figure 5: Finite Sample Power, Right Tail, $n = 250$, $\delta = -0.95$, $c = 50$



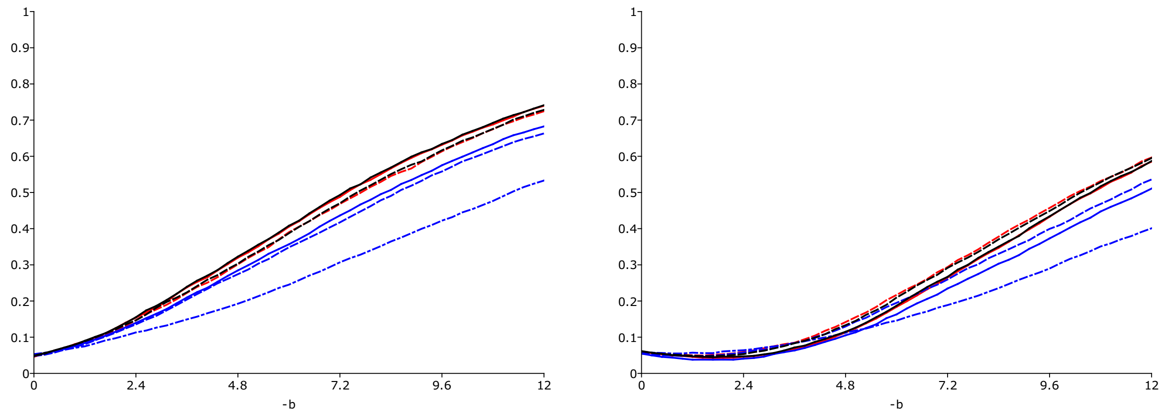
$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

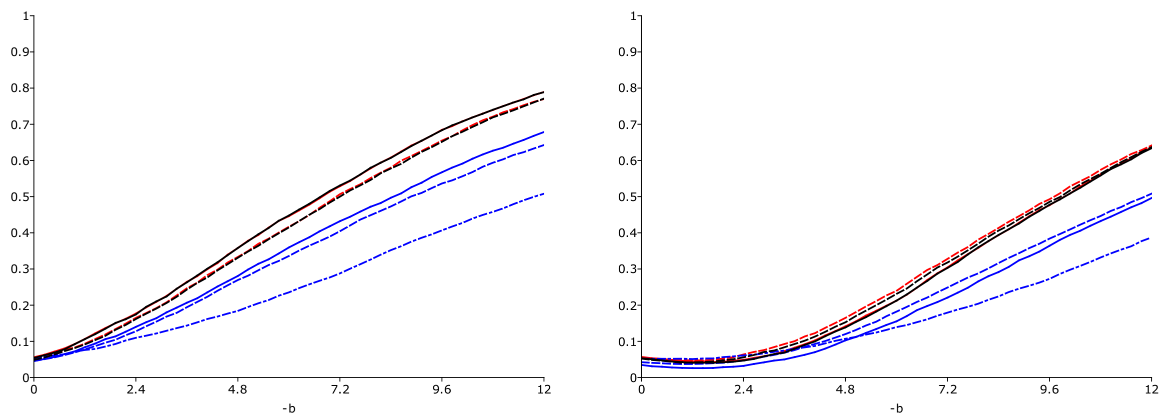
Figure 6: Finite Sample Power, Left Tail, $n = 250$, $\delta = -0.95$, $c = 2$

One Sided | Two Sided

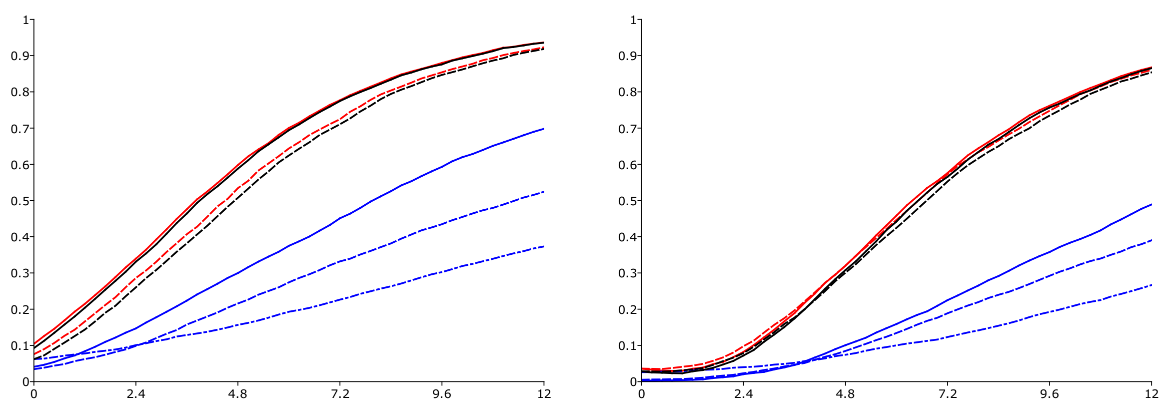
$\alpha = 0$



$\alpha = 1$



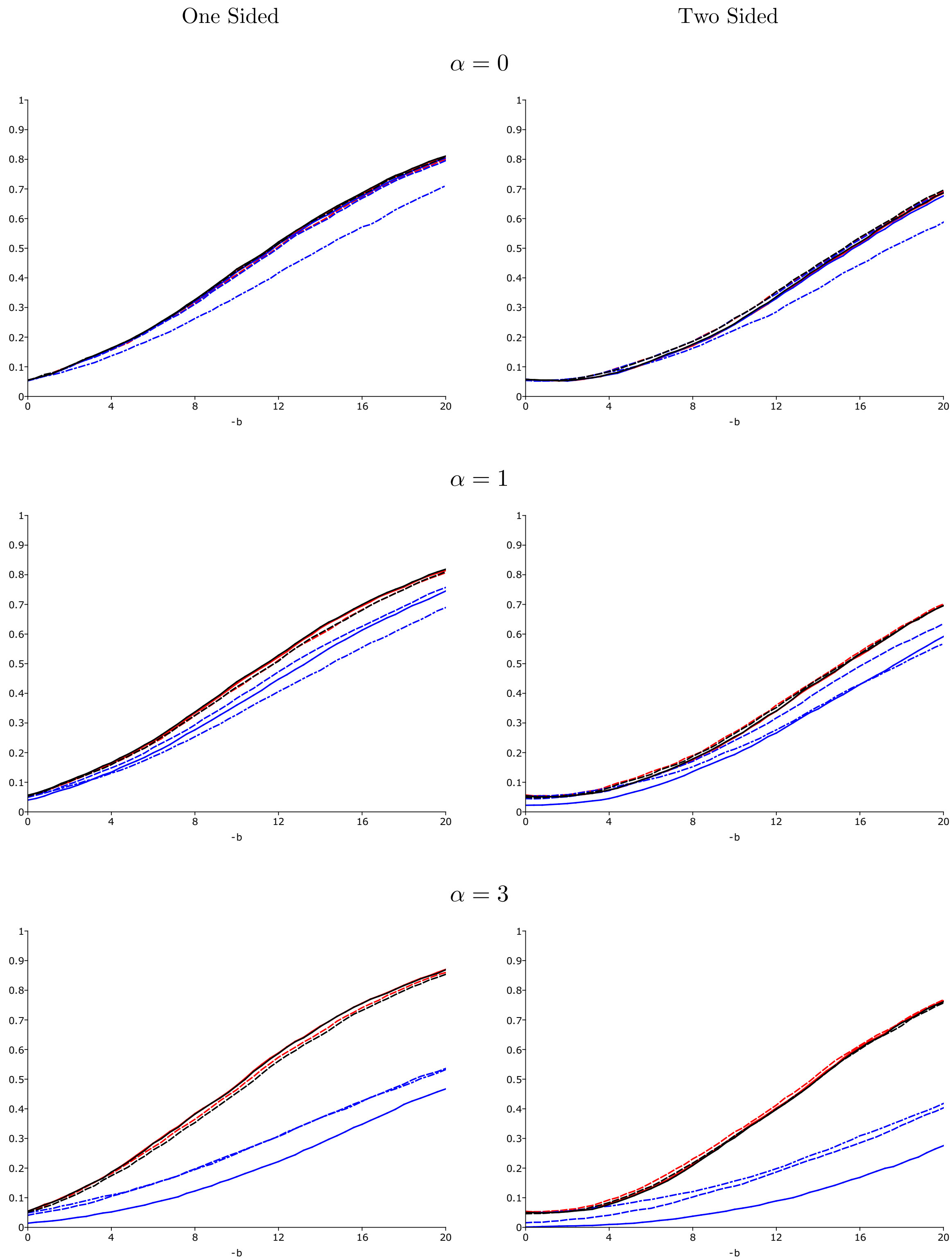
$\alpha = 3$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

Figure 7: Finite Sample Power, Left Tail, $n = 250$, $\delta = -0.95$, $c = 20$

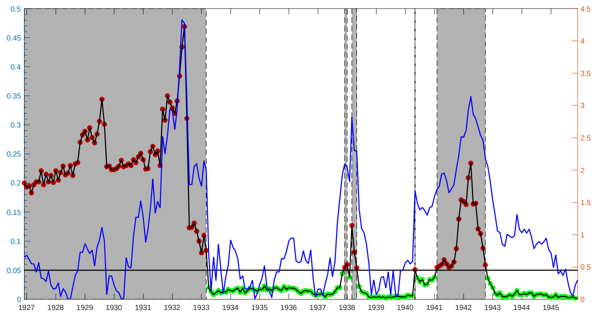


$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

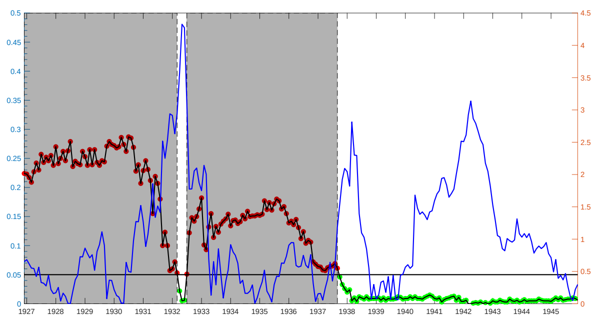
$U_{0.95}$: ——, $U_{0.75}$: – –

Figure 8: Test p-values and Estimated Magnitude of Initial Condition - Monthly CRSP 1926-1994 (Predictor $= d - p$)

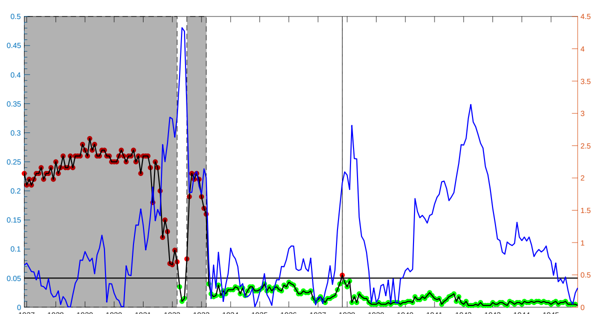
(a)$IVX_{0.95}$ (Rejection Rate 55.0%)



(b) $IVX^{\dagger}_{0.95}$ (Rejection Rate 44.5%)



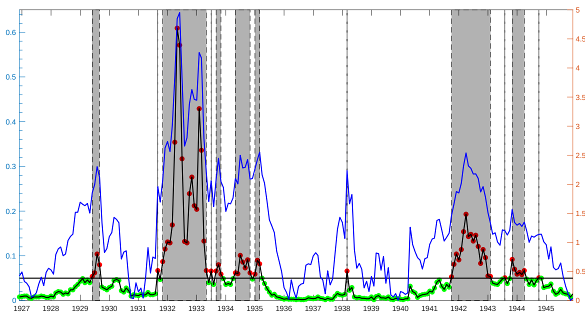
(c) $U_{0.95}$ (Rejection Rate 67.7%)



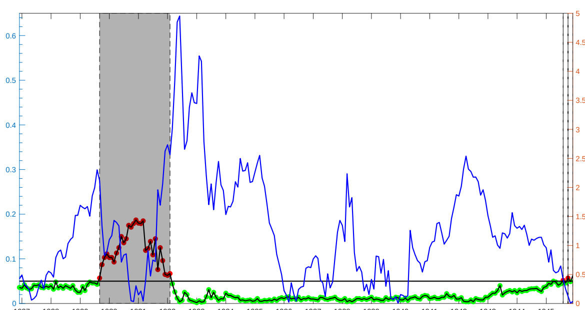
p-value: —— (Left Axis), $|\hat{\alpha}|$: —— (Right Axis)

Figure 9: Test p-values and Estimated Magnitude of Initial Condition - Monthly CRSP 1926-1994 (Predictor $= e - p$)

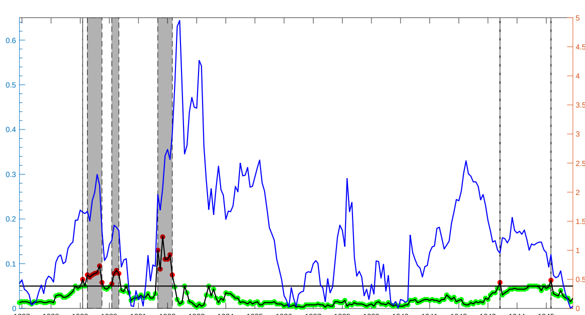
(a)$IVX_{0.95}$ (Rejection Rate 72.1%)



(b) $IVX^{\dagger}_{0.95}$ (Rejection Rate 85.6%)



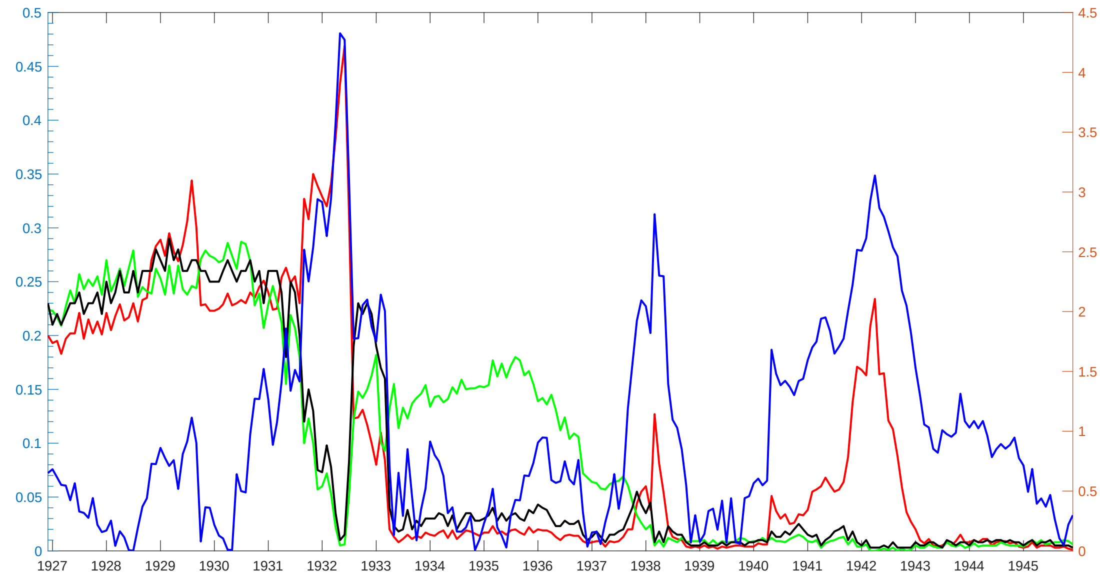
(c) $U_{0.95}$ (Rejection Rate 90.8%)



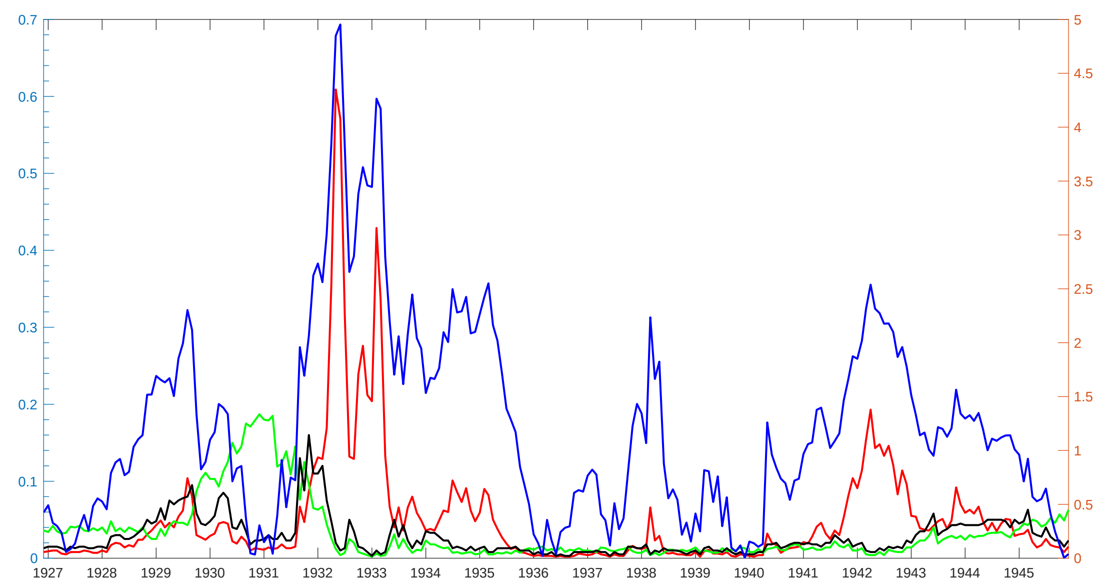
p-value: ——— (Left Axis), $|\hat{\alpha}|$: ——— (Right Axis)

Figure 10: Test p-values and Estimated Magnitude of Initial Condition - Monthly CRSP 1926-1994

(a) Predictor $= d - p$



(b) Predictor $= e - p$



$p(IVX_{0.95})$: red, $p(IVX^{\dagger}_{0.95})$: green, $p(U_{0.95})$: black (Left Axis), $|\hat{\alpha}|$: blue (Right Axis)

Supplementary Appendix to "IVX Tests for Return Predictability and the Initial Condition"

Sam Astill[a], Tassos Magdalinos[b] and A.M. Robert Taylor[a]

[a] Essex Business School, University of Essex

[b] University of Southampton

June 27, 2025

Abstract

This supplementary appendix contains three sections. Section S.1 provides proofs of the technical results stated in the main paper. Section S.2 details the implementation of the bootstrap algorithm from Demetrescu *et al.* (2022) to generate bootstrap p-values and critical values for the tests detailed in the main paper. Finally, Section S.3, provides the additional Monte Carlo results referred to in Section 5 of the main paper.

*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 Mathematical Proofs

S.1.1 Proof of Lemma 1

The proof of convergence in distribution of $[U_n(r), B_n(r), \zeta_n] \Rightarrow [U(r), B(r), \zeta]$ on $D[0,1]$ is similar to that of $[U_n(r), B_n(r), Y_n(r)]$ in Lemma 5 of Magdalinos and Petrova (2025); here the random element $Y_n(r)$ in Magdalinos and Petrova (2025) associated with a mildly explosive instrument is replaced by ζ_n associated with a (near-stationary) IVX instrument. A standard martingale approximation gives $B_n(r) = C(1)\, n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} e_t + o_p(1)$ under Assumption 1(ii), so we may write

$$[U_n(r), B_n(r), \zeta_n(r)]' = \sum_{t=1}^{\lfloor nr \rfloor} \xi_{n,t} + o_p(1) \tag{S.1}$$

where $\xi_{n,t} = \left[n^{-1/2} \left(1-\varphi_n^2\right)^{1/2} z_{t-1}\varepsilon_t, C(1)\, n^{-1/2} e_t, \left(1-\varphi_n^2\right)^{1/2} \varphi_n^{t-1} \varepsilon_t \right]'$ is a $\mathcal{F}_t$-martingale difference array and we may apply a Lindeberg-type functional CLT for vector-valued martingale difference arrays to (S.1): see Theorem 3.33 (pp. 478) of Jacod and Shiryaev (2003). The conditional Lindeberg condition on $\left\|\xi_{n,t}\right\|^2$ (3.31 in Jacod and Shiryaev (2003)) is implied by the stronger unconditional Lindeberg condition (LC) on $\left\|\xi_{n,t}\right\|^2$ which, in turn, is implied by establishing the LC on each of the components of $\xi_{n,t}$; this is done in the proof of Lemma 5 of Magdalinos and Petrova (2025) (the argument for $\zeta_n(r)$ is identical to that for $Y_n(r)$ in that paper by replacing φ_{2n}^{-1} by φ_n). The conditional variance matrix of the array in (S.1) is given by $V^{(n)} := \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{E}_{\mathcal{F}_{t-1}}\left(\xi_{n,t}\xi_{n,t}'\right)$; denoting the typical elements of $V^{(n)}$ by $\left[V_{ij}^{(n)}\right]_{i,j=1}^{3}$: $V_{11}^{(n)} \to_p \sigma^2\omega^2 r$, $V_{22}^{(n)} \to_p \omega^2 r$; $V_{12}^{(n)} \to_p 0$ by the proof of Lemma 5 in MP (2023); $V_{33}^{(n)} = \sigma^2\left(1-\varphi_n^2\right)\sum_{t=1}^{\lfloor nr \rfloor} \varphi_n^{2t} \to \sigma^2$ for all $r > 0$;

$$V_{23}^{(n)} = \omega^2 n^{-1/2}\left(1-\varphi_n^2\right)^{1/2} \sum_{t=1}^{\lfloor nr \rfloor} \varphi_n^t = O\left(\left[n\left(1-\varphi_n\right)\right]^{-1/2}\right) = o(1).$$

Finally, $V_{13}^{(n)} = \omega^2\left(1-\varphi_n^2\right) n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} \varphi_n^{t-1} z_{t-1}$ satisfies

$$\left\|V_{13}^{(n)}\right\|_{L_1} \le \sigma^2 \max_{t\le n} \left\|\left(1-\varphi_n^2\right)^{1/2} z_t\right\|_{L_2} \left(1-\varphi_n^2\right)^{1/2} n^{-1/2} \sum_{t=1}^{n} \varphi_n^t = O\left(\left(n\left(1-\varphi_n\right)\right)^{-1/2}\right).$$

We conclude that $V^{(n)} \to_p diag\left(\sigma^2\omega^2 r, \omega^2 r, \sigma^2\right)$ for $r \in [0,1]$, and applying Theorem 3.33 of Jacod and Shiryaev (2003) to (S.1), $\sum_{t=1}^{\lfloor nr \rfloor} \xi_{n,t} \Rightarrow \xi(s)$ where $\xi(r)$ is a continuous Gaussian martingale with quadratic variation $\langle \xi \rangle_r = diag\left(\sigma^2\omega^2 r, \omega^2 r, \sigma^2\right)$. By Levy's characterisation (e.g. Theorem 4.4 II of Jacod and Shiryaev (2003), $\xi(r)$ is characterised by its quadratic variation process, $\xi(r) \overset{d}{=} [U(r), B(r), \zeta]'$ with the right side defined in the statement of the lemma and independence between the components of $\xi(r)$ implied by the diagonality of the quadratic variation matrix $\langle \xi \rangle_s$.

S.1.2 Proof of Lemma 2

Using (20) we obtain

$$\sum_{t=1}^{n} \left(\tilde{z}_{t-1} - \tilde{z}_{0,t-1}\right) \varepsilon_t = \tilde{z}_0(n) \sum_{t=1}^{n} \varphi_n^{t-1} \varepsilon_t - \frac{X_0(n)(1-\rho_n)}{\varphi_n - \rho_n} \sum_{t=1}^{n} \left(\varphi_n^{t-1} - \rho_n^{t-1}\right) \varepsilon_t. \tag{S.2}$$

When $(1-\rho_n)/(1-\varphi_n) = O(1)$ and $n|\varphi_n - \rho_n| \to \infty$, the second term on the right of (S.2) is asymptotically equivalent to

$$X_0(n) \frac{1-\rho_n}{1-\varphi_n} \sum_{t=1}^{n} \rho_n^{t-1} \varepsilon_t = X_0(n) O_p\left(\frac{(1-\rho_n)^{1/2}}{1-\varphi_n}\right) = o_p\left(n^{1/2}\left(1-\varphi_n^2\right)^{-1/2}\right)$$

because $X_0(n) O_p\left((1-\rho_n)^{1/2}(1-\varphi_n)^{-1}\right)$ is $O_p\left(n^{1/2}\right) o_p\left((1-\varphi_n)^{-1/2}\right)$ when $c \in \mathbb{R}$ and $o_p\left(n^{1/2}\right) O_p\left((1-\varphi_n)^{-1/2}\right)$ when $c = -\infty$. Since $n^{-1/2}\left(1-\rho_n^2\varphi_n^2\right)^{1/2} \sim n^{-1/2}\left(1-\varphi_n^2\right)^{1/2}$ when $(1-\rho_n)/(1-\varphi_n) = O(1)$, this proves part (i) in this case. When $n|\varphi_n - \rho_n| = O(1)$, the mean value theorem implies that there exists ϕ_n satisfying $|\varphi_n - \phi_n| < |\varphi_n - \rho_n|$ and

$$\sum_{t=1}^{n} \frac{\varphi_n^{t-1} - \rho_n^{t-1}}{\varphi_n - \rho_n} \varepsilon_t = \sum_{t=1}^{n} (t-1) \phi_n^{t-1} \varepsilon_t = O_p\left(\sum_{t=1}^{n} t^2 \phi_n^{2t}\right)^{1/2} = O_p\left((1-\phi_n)^{-3/2}\right) = O_p\left(\kappa_n^{3/2}\right)$$

from the asymptotic equivalence

$$\sum_{t=1}^{n} t^p \varphi_n^{2t} \sim (1-\varphi_n)^{-1-p} \frac{\Gamma(p+1)}{2^{p+1}} \text{ as } n \to \infty. \tag{S.3}$$

We conclude that the second term on the right of (S.2) is of order

$$X_0(n) O_p\left((\rho_n - 1)\kappa_n^{3/2}\right) = o_p\left(n^{1/2}\right) O_p\left(\kappa_n^{1/2}\right) = o_p\left(n^{1/2}\left(1-\rho_n^2\varphi_n^2\right)^{-1/2}\right)$$

since $\left(1-\rho_n^2\varphi_n^2\right)^{-1} \sim \kappa_n$ when $n|\varphi_n - \rho_n| = O(1)$. We conclude that the second term on the right of (S.2) is $o_p\left(n^{1/2}\left(1-\rho_n^2\varphi_n^2\right)^{-1/2}\right)$ when $(1-\rho_n)/(1-\varphi_n) = O(1)$; when

$c=-\infty$, the same holds for the first term:

$$\tilde{z}_0(n)\sum_{t=1}^{n}\varphi_n^{t-1}\varepsilon_t = o_p\left(n^{1/2}\right)O_p\left(\left(1-\varphi_n^2\right)^{-1/2}\right) = o_p\left(n^{1/2}\left(1-\rho_n^2\varphi_n^2\right)^{-1/2}\right).$$

Next, we show that the right of (S.2) is $o_p\left(n^{1/2}\left(1-\rho_n^2\varphi_n^2\right)^{-1/2}\right)$ when $(1-\varphi_n)/(1-\rho_n)\to 0$. Since

$$\begin{aligned}\frac{X_0(n)(\rho_n-1)}{\varphi_n-\rho_n}\sum_{t=1}^{n}\left(\varphi_n^{t-1}-\rho_n^{t-1}\right)\varepsilon_t &= -X_0(n)\left[1+O\left(\frac{1-\varphi_n}{1-\rho_n}\right)\right]\sum_{t=1}^{n}\left(\varphi_n^{t-1}-\rho_n^{t-1}\right)\varepsilon_t\\ &= -X_0(n)\sum_{t=1}^{n}\varphi_n^{t-1}\varepsilon_t + X_0(n)O_p\left(\kappa_n^{1/2}\right)\end{aligned}$$

(S.2) implies that

$$\begin{aligned}\sum_{t=1}^{n}\left(\tilde{z}_{t-1}-\tilde{z}_{0,t-1}\right)\varepsilon_t &= \left(\tilde{z}_0(n)-X_0(n)\right)\sum_{t=1}^{n}\varphi_n^{t-1}\varepsilon_t + X_0(n)O_p\left(\kappa_n^{1/2}\right)\\ &= \left(\tilde{z}_0(n)-X_0(n)\right)O_p\left[\left(1-\varphi_n^2\right)^{-1/2}\right] + o_p\left(n^{1/2}\kappa_n^{1/2}\right) = o_p\left(n^{1/2}\kappa_n^{1/2}\right)\end{aligned}$$

by Assumption 2(c), thereby completing the proof of part (i) of the lemma.

For part (ii), (19) gives

$$\underline{x}_{t-1} = x_{t-1}-\bar{x}_{n-1} = x_{0,t-1}-\bar{x}_{0,n-1}+X_0(n)\left(\rho_n^{t-1}-\frac{1}{n}\frac{1-\rho_n^n}{1-\rho_n}\right).$$

Combining the above with (20), we obtain

$$\begin{aligned}R_{1n} &= \sum_{t=1}^{n}\left(x_{t-1}-\bar{x}_{n-1}\right)\tilde{z}_{t-1}-\sum_{t=1}^{n}\left(x_{0,t-1}-\bar{x}_{0,n-1}\right)\tilde{z}_{0,t-1}\\
&= \tilde{z}_0(n)\sum_{t=1}^{n}\varphi_n^{t-1}\left(x_{0,t-1}-\bar{x}_{0,n-1}\right)\\
&\quad +X_0(n)\left\{(1-\rho_n)\sum_{t=1}^{n}\left(x_{0,t-1}-\bar{x}_{0,n-1}\right)\frac{\varphi_n^{t-1}-\rho_n^{t-1}}{\varphi_n-\rho_n}+\sum_{t=1}^{n}\left(\rho_n^{t-1}-\frac{1}{n}\frac{1-\rho_n^n}{1-\rho_n}\right)\tilde{z}_{t-1}\right\}\\
&= \tilde{z}_0(n)\sum_{t=1}^{n}\varphi_n^{t-1}x_{0,t-1}-\bar{x}_{0,n-1}\tilde{z}_0(n)\sum_{t=1}^{n}\varphi_n^{t-1}\\
&\quad +X_0(n)(1-\rho_n)\sum_{t=1}^{n}\left(x_{0,t-1}-\bar{x}_{0,n-1}\right)\frac{\varphi_n^{t-1}-\rho_n^{t-1}}{\varphi_n-\rho_n}\\
&\quad +X_0(n)\sum_{t=1}^{n}\left(\rho_n^{t-1}-\frac{1}{n}\frac{1-\rho_n^n}{1-\rho_n}\right)\tilde{z}_{0,t-1}+X_0(n)\tilde{z}_0(n)\sum_{t=1}^{n}\left(\rho_n^{t-1}-\frac{1}{n}\frac{1-\rho_n^n}{1-\rho_n}\right)\varphi_n^{t-1}\\
&\quad +X_0(n)^2(1-\rho_n)\sum_{t=1}^{n}\left(\rho_n^{t-1}-\frac{1}{n}\frac{1-\rho_n^n}{1-\rho_n}\right)\frac{\varphi_n^{t-1}-\rho_n^{t-1}}{\varphi_n-\rho_n}. \qquad (S.4)\end{aligned}$$

Since

$$\sum_{t=1}^{n-1} \varphi_n^t x_{0,t} = O_p\left(\left(1-\varphi_n^2\right)^{-1/2}\left(1-\rho_n\varphi_n\right)^{-1}\right),$$

the first term of (S.4) is $O_p\left(n^{1/2}\left(1-\varphi_n^2\right)^{-1/2}\left(1-\rho_n\varphi_n\right)^{-1}\right) = o_p\left(n\left(1-\rho_n\varphi_n\right)^{-1}\right)$. The remaining terms of (S.4) will be $o_p\left(n\left(1-\rho_n\varphi_n\right)^{-1}\right)$ only if $\kappa_n/n \to 0$: the second term of (S.4) satisfies $\bar{x}_{0,n-1}\tilde{z}_0\left(n\right)\left(1-\varphi_n\right)^{-1} = \tilde{z}_0\left(n\right)O_p\left(n^{-1/2}\kappa_n\left(1-\varphi_n\right)^{-1}\right) = o_p\left(\kappa_n\left(1-\varphi_n\right)^{-1}\right)$ when $\kappa_n/n \to 0$; the third term has order $X_0\left(n\right)O_p\left(\left(1-\varphi_n\rho_n\right)^{-1}\left[\kappa_n^{1/2} \vee \left(1-\varphi_n\right)^{-1/2}\right]\right) = o_p\left(n\left(1-\rho_n\varphi_n\right)^{-1}\right)$ when $\kappa_n/n \to 0$; since $\sum_{t=1}^{n-1}\rho_n^t\tilde{z}_t = O_p\left(\kappa_n^{1/2}\left(1-\varphi_n\rho_n\right)^{-1}\right)$, the fourth term has order $o_p\left(n^{1/2}\kappa_n^{1/2}\left(1-\varphi_n\rho_n\right)^{-1}\right)$; the fifth and sixth terms have order $o_p\left(n\left(1-\varphi_n\rho_n\right)^{-1}\right)$ when $X_0\left(n\right) = o_p\left(n^{1/2}\right)$ (since $\kappa_n/n \to 0$). We conclude that, when $\kappa_n/n \to 0$,

$$\left(1-\varphi_n^2\rho_n^2\right)\frac{1}{n}R_{1n} = o_p\left(1\right).$$

In the local-to-unit root case, all but the first term of (S.4) contribute asymptotically and we obtain

$$\left(1-\varphi_n^2\rho_n^2\right)\frac{1}{n}R_{1n} = 2\left(1+o\left(1\right)\right)\left(1-\varphi_n\right)\frac{1}{n}R_{1n}$$

where

$$\begin{aligned}
\left(1-\varphi_n\right)\frac{1}{n}R_{1n} &= -\frac{\bar{x}_{0,n-1}}{\sqrt{n}}\frac{\tilde{z}_0\left(n\right)}{\sqrt{n}}\left(1-\varphi_n\right)\sum_{t=1}^{n}\varphi_n^{t-1} \\
&\quad + X_0\left(n\right)\left(1-\rho_n\right)\frac{1}{n}\left(\sum_{t=1}^{n}\rho_n^{t-1}x_{0,t-1} - \bar{x}_{0,n-1}\sum_{t=1}^{n}\rho_n^{t-1}\right) \\
&\quad + \frac{X_0\left(n\right)}{\sqrt{n}}\left(1-\varphi_n\right)\frac{1}{\sqrt{n}}\left(\sum_{t=1}^{n}\rho_n^{t-1}z_{0,t-1} + \frac{1-e^c}{c}\sum_{t=1}^{n}z_{0,t-1}\right) \\
&\quad + \frac{1}{n}X_0\left(n\right)\tilde{z}_0\left(n\right)\left(1+\frac{1-e^c}{c}\right)\left(1-\varphi_n\right)\sum_{t=1}^{n}\varphi_n^{t-1} \\
&\quad + \frac{1}{n}X_0\left(n\right)^2\left(1-\rho_n\right)\left(\sum_{t=1}^{n}\rho_n^{2(t-1)} + \frac{1-e^c}{c}\sum_{t=1}^{n}\rho_n^{t-1}\right)
\end{aligned}$$

with the convention $\frac{1-e^c}{c} := \lim_{c\to 0}\frac{1-e^c}{c} = 0$ when $c=0$. Since

$$\begin{aligned}
(1-\varphi_n)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\rho_n^{t-1}z_{0,t-1} &= (1-\varphi_n)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\rho_n^{t-1}\sum_{j=1}^{t-1}\varphi_n^{t-1-j}u_j \\
&= (1-\varphi_n)\frac{1}{\sqrt{n}}\sum_{j=1}^{n-1}\varphi_n^{-j}u_j\sum_{t=j+1}^{n}(\rho_n\varphi_n)^{t-1} \\
&= (1-\varphi_n)\frac{1}{\sqrt{n}}\sum_{j=1}^{n-1}\rho_n^{j}u_j\frac{1-(\rho_n\varphi_n)^{n-j}}{1-\rho_n\varphi_n} \\
&= \frac{1-\varphi_n}{1-\rho_n\varphi_n}\frac{1}{\sqrt{n}}\left[\sum_{j=1}^{n-1}\rho_n^{j}u_j-\rho_n^{n}\sum_{j=1}^{n-1}\varphi_n^{n-j}u_j\right] \\
&= [1+o_p(1)]\frac{1}{\sqrt{n}}\sum_{j=1}^{n-1}\rho_n^{j}u_j
\end{aligned}$$

because $\left\|n^{-1/2}\sum_{j=1}^{n-1}\varphi_n^{n-j}u_j\right\|_{L_2} = O\left(n^{-1}(1-\varphi_n)^{-1}\right)$ and $n^{-1/2}(1-\varphi_n)\sum_{t=1}^{n}z_{0,t-1} = n^{-1/2}\sum_{t=1}^{n}u_t + o_p(1)$, we conclude that

$$\left(1-\varphi_n^2\rho_n^2\right)\frac{1}{n}R_{1n} = -2\frac{\tilde{z}_0(n)}{\sqrt{n}}\frac{1}{n^{3/2}}\sum_{t=1}^{n}x_{0,t-1}+2\frac{X_0(n)}{\sqrt{n}}\psi_n(c)+o_p(1)$$

where $\psi_n(c)$ is given in (21).

For part (iii), (20) gives

$$\begin{aligned}
\sum_{t=1}^{n}\tilde{z}_t^2 &= \sum_{t=1}^{n}\left(\tilde{z}_{0,t}+\varphi_n^t\tilde{z}_0(n)\right)^2+X_0(n)^2(1-\rho_n)^2\sum_{t=1}^{n}\left(\frac{\varphi_n^t-\rho_n^t}{\varphi_n-\rho_n}\right)^2 \\
&\quad -2X_0(n)(1-\rho_n)\sum_{t=1}^{n}\left(\tilde{z}_{0,t}+\varphi_n^t\tilde{z}_0(n)\right)\frac{\varphi_n^t-\rho_n^t}{\varphi_n-\rho_n}. \qquad \text{(S.5)}
\end{aligned}$$

When $(1-\rho_n)/(1-\varphi_n)=O(1)$, the second term of (S.5) of has order

$$X_0(n)^2 O_p\left(\kappa_n^{-1}(1-\rho_n\varphi_n)^{-2}\right) = X_0(n)^2 O_p\left(\kappa_n^{-1}(1-\rho_n\varphi_n)^{-2}\right) = o_p\left(n(1-\rho_n\varphi_n)^{-1}\right)$$

by direct estimation when $n|\varphi_n-\rho_n|\to\infty$ and by applying the mean value theorem to $(\varphi_n^t-\rho_n^t)/(\varphi_n-\rho_n)$ when $n|\varphi_n-\rho_n|=O(1)$; the order $o_p\left(n(1-\rho_n\varphi_n)^{-1}\right)$ applies both when $\kappa_n/n\to 0$ (since $\kappa_n^{-1}(1-\rho_n\varphi_n)^{-2}=O\left((1-\rho_n\varphi_n)^{-1}\right)$ and $X_0(n)^2=o_p(n)$) and when $\kappa_n=n$ (since $\kappa_n^{-1}(1-\rho_n\varphi_n)^{-2}=o\left((1-\rho_n\varphi_n)^{-1}\right)$). Similarly, $\sum_{t=1}^{n-1}\rho_n^t\tilde{z}_{0,t} =$

$O_p\left(\kappa_n^{1/2}\left(1-\varphi_n\rho_n\right)^{-1}\right)$ implies that when $\left(1-\rho_n\right)/\left(1-\varphi_n\right)=O\left(1\right)$, the third term of (S.5) of has order

$$X_0\left(n\right)O_p\left(\kappa_n^{-1/2}\left(1-\varphi_n\rho_n\right)^{-2}\right)+X_0\left(n\right)\tilde{z}_0\left(n\right)O_p\left(\kappa_n^{-1}\left(1-\rho_n\varphi_n\right)^{-2}\right)=o_p\left(n\left(1-\varphi_n\rho_n\right)^{-1}\right)$$

because $X_0\left(n\right)=o_p\left(n^{1/2}\right)$ when $\kappa_n/n\to 0$ and $\kappa_n^{-1}\left(1-\varphi_n\rho_n\right)^{-1}\to 0$ when $\kappa_n=n$. We conclude that, when $\left(1-\rho_n\right)/\left(1-\varphi_n\right)=O\left(1\right)$,

$$\begin{aligned}\frac{1}{n}\left(1-\rho_n\varphi_n\right)\sum_{t=1}^{n}\left(\tilde{z}_t^2-\tilde{z}_{0,t}^2\right) &= \tilde{z}_0\left(n\right)^2\frac{1}{n}\left(1-\rho_n\varphi_n\right)\sum_{t=1}^{n}\varphi_n^{2t}+2\tilde{z}_0\left(n\right)\frac{1}{n}\left(1-\rho_n\varphi_n\right)\sum_{t=1}^{n}\tilde{z}_{0,t}\varphi_n^t\\ &= \frac{1}{2}\frac{\tilde{z}_0\left(n\right)^2}{n}\left[1+O\left(\kappa_n^{-1}\left(1-\varphi_n\right)^{-1}\right)\right]+o_p\left(1\right) \qquad \text{(S.6)}\end{aligned}$$

because $n^{-1/2}\left(1-\rho_n\varphi_n\right)\sum_{t=1}^{n}\tilde{z}_{0,t}\varphi_n^t=O_p\left(n^{-1/2}\left(1-\rho_n\varphi_n\right)^{-1/2}\right)$. Note that by Assumption 2, the right side of (S.6) is $o_p\left(1\right)$ when $\kappa_n/n\to 0$.

It remains to show that, when $\left(1-\varphi_n\right)/\left(1-\rho_n\right)\to 0$,

$$\frac{1}{n\kappa_n}\sum_{t=1}^{n}\tilde{z}_t^2-\frac{1}{n\kappa_n}\sum_{t=1}^{n}\tilde{z}_{0,t}^2=o_p\left(1\right), \qquad \text{(S.7)}$$

since in this case, $1-\rho_n\varphi_n\sim\kappa_n^{-1}$. Rearranging (S.5), we obtain

$$\begin{aligned}\sum_{t=1}^{n}\tilde{z}_t^2-\sum_{t=1}^{n}\tilde{z}_{0,t}^2 &= 2\left[\tilde{z}_0\left(n\right)-X_0\left(n\right)\right]\sum_{t=1}^{n}\varphi_n^t\tilde{z}_{0,t}+\left[\tilde{z}_0\left(n\right)-X_0\left(n\right)\right]^2\sum_{t=1}^{n}\varphi_n^{2t}\\ &\quad+\left[X_0\left(n\right)^2\left(\left(\frac{1-\rho_n}{\varphi_n-\rho_n}\right)^2-1\right)-2X_0\left(n\right)\tilde{z}_0\left(n\right)\left(\frac{1-\rho_n}{\varphi_n-\rho_n}-1\right)\right]\sum_{t=1}^{n}\varphi_n^{2t}\\ &\quad-2X_0\left(n\right)\left(\frac{1-\rho_n}{\varphi_n-\rho_n}-1\right)\sum_{t=1}^{n}\tilde{z}_{0,t}\varphi_n^t+2X_0\left(n\right)\frac{1-\rho_n}{\varphi_n-\rho_n}\sum_{t=1}^{n}\tilde{z}_{0,t}\rho_n^t\\ &\quad+2X_0\left(n\right)\tilde{z}_0\left(n\right)\frac{1-\rho_n}{\varphi_n-\rho_n}O\left(\kappa_n\right)+X_0\left(n\right)^2\left(\frac{1-\rho_n}{\varphi_n-\rho_n}\right)^2O\left(\kappa_n\right). \qquad \text{(S.8)}\end{aligned}$$

When $\left(1-\varphi_n\right)/\left(1-\rho_n\right)\to 0$, $\sum_{t=1}^{n-1}\varphi_n^t\tilde{z}_{0,t}=O_p\left(\left(1-\varphi_n^2\right)^{-1/2}\kappa_n\right)$ and Assumption 2(c) implies that the first term on the right of (S.8) has order

$$o_p\left(n^{1/2}\kappa_n^{1/2}\left(1-\varphi_n\right)^{1/2}\right)O_p\left(\left(1-\varphi_n^2\right)^{-1/2}\kappa_n\right)=o_p\left(n^{1/2}\kappa_n^{3/2}\right)=o_p\left(n\kappa_n\right).$$

Similarly, the second term on the right of (S.8) has order $o_p\left(n\kappa_n\right)$. Since $\left(1-\varphi_n\right)/\left(1-\rho_n\right)\to 0$ implies that both $\frac{1-\rho_n}{\varphi_n-\rho_n}-1$ and $\left(\frac{1-\rho_n}{\varphi_n-\rho_n}\right)^2-1$ are $O\left[\left(1-\varphi_n\right)\kappa_n\right]$ and $X_0\left(n\right)$ and $\tilde{z}_0\left(n\right)$

are both $o_p\left(n^{1/2}\right)$, the third and fourth terms on the right of (S.8) have order $o_p\left(n\kappa_n\right)$. The fifth term has order $o_p\left[n^{1/2}\left(1-\varphi_n\right)^{1/2}\kappa_n^2\right] = o_p\left(n^{1/2}\kappa_n^{3/2}\right) = o_p\left(n\kappa_n\right)$. For the last three terms on the right of (S.8), $\left(1-\rho_n\right)/\left(\varphi_n-\rho_n\right) = O\left(1\right)$, so the sixth term has order $o_p\left[n^{1/2}\left(1-\varphi_n\right)^{1/2}\kappa_n^2\right] = o_p\left(n^{1/2}\kappa_n^{3/2}\right) = o_p\left(n\kappa_n\right)$, and the seventh and eighth terms are $o_p\left(n\kappa_n\right)$. This establishes (S.7) and completes the proof of part (iii) when $c=-\infty$. When $c\in\mathbb{R}$, $1-\rho_n^2\varphi_n^2 \sim 2(1-\rho_n\varphi_n)$ so (S.6) implies that

$$\frac{1}{n}\left(1-\rho_n^2\varphi_n^2\right)\left(\sum_{t=0}^{n-1}\tilde{z}_t^2-\sum_{t=0}^{n-1}\tilde{z}_{0,t}^2\right) = \left(\frac{\tilde{z}_0\left(n\right)}{\sqrt{n}}\right)^2+o_p\left(1\right)$$

completing the proof of part (iii).

S.1.3 Proof of Theorem 1

Firstly, we show that, as in the standard case of a zero instrument initialisation, the KMS correction does not affect the limit distribution of the t-statistic; we need to show that

$$\left(1-\rho_n^2\varphi_n^2\right)\bar{z}_{n-1}^2 = o_p\left(1\right). \tag{S.9}$$

By (20),

$$\begin{aligned}\left(1-\rho_n^2\varphi_n^2\right)^{1/2}\bar{z}_{n-1} &= \frac{1}{n}\left(1-\rho_n^2\varphi_n^2\right)^{1/2}\sum_{t=1}^{n}\tilde{z}_{0,t-1} &\text{(S.10)}\\ &\quad+\left[\tilde{z}_0\left(n\right)-X_0\left(n\right)\frac{1-\rho_n}{\varphi_n-\rho_n}\right]\left(1-\rho_n^2\varphi_n^2\right)^{1/2}\frac{1}{n}\sum_{t=0}^{n-1}\varphi_n^t\\ &\quad+X_0\left(n\right)\frac{1-\rho_n}{\varphi_n-\rho_n}\left(1-\rho_n^2\varphi_n^2\right)^{1/2}\frac{1}{n}\sum_{t=0}^{n-1}\rho_n^t. &\text{(S.11)}\end{aligned}$$

Since $\sum_{t=1}^{n}\tilde{z}_{0,t-1} = O_p\left(n^{1/2}\left(\kappa_n\wedge\left(1-\varphi_n\right)^{-1}\right)\right)$, the first term on the right of (S.11) is $O_p\left(n^{-1/2}\left(1-\varphi_n\right)^{-1/2}\right) = o_p\left(1\right)$. Since $X_0\left(n\right) = O_p\left(n^{1/2}\right)$, the last term of (S.11) is $O_p\left(n^{-1}X_0\left(n\right)\left(1-\rho_n^2\varphi_n^2\right)^{-1/2}\right) = O_p\left(n^{-1/2}\left(1-\rho_n^2\varphi_n^2\right)^{-1/2}\right) = o_p\left(1\right)$. When $\left(1-\rho_n\right)/\left(1-\varphi_n\right) = O\left(1\right)$, the second term of (S.11) is

$$O_p\left(n^{1/2}\right)\left(1-\varphi_n^2\right)^{1/2}\frac{1}{n}\sum_{t=0}^{n-1}\varphi_n^t = O_p\left(n^{-1/2}\left(1-\varphi_n^2\right)^{-1/2}\right) = o_p\left(1\right).$$

When $(1-\rho_n)/(1-\varphi_n) \to \infty$, the second term of (S.11) is

$$\begin{aligned}\left[\tilde{z}_0(n) - X_0(n) + O_p\left(n^{1/2}\frac{1-\varphi_n}{1-\rho_n}\right)\right] O_p\left[\kappa_n^{-1/2}\frac{1}{n}(1-\varphi_n)^{-1}\right] &= o_p\left(n^{-1/2}(1-\varphi_n)^{-1/2}\right) \\ &\quad + O_p\left(n^{-1/2}\kappa_n^{1/2}\right) \\ &= o_p(1)\end{aligned}$$

by Assumption 2(iii) on the order of $|\tilde{z}_0(n) - X_0(n)|$. This completes the proof of (S.9).

By (S.9), we conclude that $T_n^{\dagger}(\varphi_n) = \tilde{T}_n^{\dagger}(\varphi_n) + o_p(1)$, and $T_n(\varphi_n) = \tilde{T}_n(\varphi_n) + o_p(1)$ where

$$\tilde{T}_n^{\dagger}(\varphi_n) := \frac{\left|\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1}\right|}{\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1}}\hat{\sigma}_{\varepsilon}^{-1}\left(\sum_{t=1}^n \tilde{z}_{t-1}^2\right)^{-1/2}\sum_{t=1}^n \tilde{z}_{t-1}\varepsilon_t \tag{S.12}$$

where the instrument process $\tilde{z}_t$ has initialisation $\tilde{z}_0(n)$ satisfying parts (ii) and (iii) of Assumption 2 and

$$\tilde{T}_n(\varphi_n) := \frac{\left|\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1}^0\right|}{\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1}^0}\hat{\sigma}_{\varepsilon}^{-1}\left(\sum_{t=1}^n \left(\tilde{z}_{t-1}^0\right)^2\right)^{-1/2}\sum_{t=1}^n \tilde{z}_{t-1}^0\varepsilon_t \tag{S.13}$$

where $\tilde{z}_t^0 = \sum_{j=1}^t \varphi_n^{t-j}\Delta x_j$ is the restriction of $\tilde{z}_t$ with $\tilde{z}_0(n) = 0$. Lemma 2, implies that, when $c = -\infty$, Lemma 2 implies that, when $c = -\infty$, $\tilde{T}_n(\varphi_n) = T_n^0(\varphi_n) + o_p(1)$ and $\tilde{T}_n^{\dagger}(\varphi_n) = T_n^0(\varphi_n) + o_p(1)$ where

$$T_n^0(\varphi_n) = \frac{\left|\sum_{j=1}^n x_{0,j-1}\tilde{z}_{0,j-1}\right|}{\sum_{j=1}^n x_{0,j-1}\tilde{z}_{0,j-1}}\hat{\sigma}_{\varepsilon}^{-1}\left(\sum_{t=1}^n \tilde{z}_{0,t-1}^2\right)^{-1/2}\sum_{t=1}^n \tilde{z}_{0,t-1}\varepsilon_t$$

with $x_{0,t}$ and $\tilde{z}_{0,t}$ defined in (18). By Theorem 1 of Magdalinos and Petrova (2025), $T_n^0(\varphi_n) \xrightarrow{d} \zeta \overset{d}{=} N(0,1)$. We conclude that, when $c = -\infty$,

$$\left[T_n(\varphi_n), T_n^{\dagger}(\varphi_n)\right] = [1,1]\,T_n^0(\varphi_n) + o_p(1) \xrightarrow{d} [1,1]\,\zeta, \tag{S.14}$$

showing the theorem for the weakly persistent case.

For the local-to-unit root case, $c \in \mathbb{R}$, the computations

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n-1}\rho_n^j u_j = \frac{1}{\sqrt{n}}\int_1^{n-1}\rho_n^{\lfloor t\rfloor}dS_{\lfloor t\rfloor} = \int_{1/n}^{(n-1)/n}\rho_n^{\lfloor nt\rfloor}d\left(\frac{S_{\lfloor nt\rfloor}}{\sqrt{n}}\right) \xrightarrow{d} \int_0^1 e^{ct}dB(t)$$

$$\frac{1}{n^{3/2}}\sum_{t=1}^n \rho_n^{t-1}x_{0,t-1} = \frac{1}{n^{3/2}}\int_0^{n-1}\rho_n^{\lfloor t\rfloor}x_{0\lfloor t\rfloor}dt = \int_0^{(n-1)/n}\rho_n^{\lfloor nt\rfloor}\frac{x_{0\lfloor nt\rfloor}}{n^{1/2}}dt \xrightarrow{d} \int_0^1 e^{ct}J_c(t)\,dt$$

and (21) imply that $\psi_n(c) \xrightarrow{d} \psi(c)$, where

$$\psi(c) = \int_0^1 e^{ct} dB(t) + \frac{1-e^c}{c} B(1) - c\int_0^1 e^{ct} J_c(t)\,dt + (1-e^c)\int_0^1 J_c(t)\,dt$$
$$+ G_c(B, \mathbb{X}_0)\left(1 + \frac{1-e^c}{c}\right) + \mathbb{X}_0(1-e^c)\left(\frac{1}{2}(1+e^c) + \frac{1}{c}(1-e^c)\right). \quad \text{(S.15)}$$

Lemma 2(ii), (S.15) and Lemma 3 of Magdalinos and Petrova (2025) imply that

$$n^{-1}\left(1-\rho_n^2\varphi_n^2\right)\sum_{t=1}^n (x_t - \bar{x}_n)\tilde{z}_t \xrightarrow{d} \Psi_c := \omega^2 + J_c(1)^2 + 2\mathbb{X}_0\psi(c) - 2\left(J_c(1) + G_c(B, \mathbb{X}_0)\right)\int_0^1 J_c(r)\,dr. \quad \text{(S.16)}$$

Lemma 2(iii) and Lemma 3 of Magdalinos and Petrova (2025) imply

$$n^{-1}\left(1-\rho_n^2\varphi_n^2\right)\sum_{t=1}^n \tilde{z}_{t-1}^2 = \omega^2 + G_c\left(B_n, n^{-1/2}X_0(n)\right)^2 + o_p(1). \quad \text{(S.17)}$$

By standard IVX asymptotics (Phillips and Magdalinos (2009)), $n^{-1/2}\left(1-\rho_n^2\varphi_n^2\right)^{1/2}\sum_{t=1}^n \tilde{z}_{0,t-1}\varepsilon_t = U_n(1) + o_p(1)$ when $c \in \mathbb{R}$; hence Lemma 2(i) gives

$$n^{-1/2}\left(1-\varphi_n^2\right)^{1/2}\sum_{t=1}^n \tilde{z}_{t-1}\varepsilon_t = U_n(1) + n^{-1/2}\tilde{z}_0(n)\zeta_n + o_p(1)$$
$$= \left[1, G_c\left(B_n, n^{-1/2}X_0(n)\right)\right]\left[U_n(1), \zeta_n(1)\right]' + o_p(1) \quad \text{(S.18)}$$

Using (S.17) and (S.18), the t-statistic satisfies

$$\tilde{T}_n^\dagger(\varphi_n) = [1 + o_p(1)]\frac{\left|\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1}\right|}{\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1}} \times \frac{1}{\hat{\sigma}_\varepsilon}\left(\frac{1}{n}\left(1-\varphi_n^2\right)\sum_{t=1}^n \tilde{z}_{t-1}^2\right)^{-1/2}\frac{1}{\sqrt{n}}\left(1-\varphi_n^2\right)^{1/2}\sum_{t=1}^n \tilde{z}_{t-1}\varepsilon_t$$
$$= [1 + o_p(1)]\frac{\left|\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1}\right|}{\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1}}\left(\omega^2 + G_c\left(B_n, n^{-1/2}X_0(n)\right)^2\right)^{-1/2}\left[\omega, G_c\left(B_n, n^{-1/2}X_0(n)\right)\right] w_n$$

where, by Lemma 1,

$$w_n := \frac{1}{\sigma_\varepsilon}\left[\frac{1}{\omega}U_n(1), \zeta_n(1)\right]' \xrightarrow{d} w \overset{d}{=} N(0, I_2). \quad \text{(S.19)}$$

By Lemma 2(ii), $n^{-1}\left(1-\varphi_n^2\right)\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1} = g_c\left(B_n, n^{-1/2}X_0(n)\right) + o_p(1)$, where g_c is a continuous function on $D[0,1] \times \mathbb{R}$ and the function $x \mapsto sign(x)$ is $\mathbb{P}_{B,X_0}$-$a.s.$ continuous

($\mathbb{P}_{B,X_0}(0) = 0$ as $\mathbb{P}_{B,X_0}$ is a Gaussian measure by Assumption 2(i)), (S.16), (S.17) and the continuous mapping theorem imply that

$$\tilde{T}_n^{\dagger}(\varphi_n) \xrightarrow{d} sign(\Psi_c)\left(\omega^2 + G_c(B, \mathbb{X}_0)^2\right)^{-1/2}[\omega, G_c(B, \mathbb{X}_0)]\, w = \left(\frac{\upsilon_c}{\|\upsilon_c\|}\right)' w$$

where $\upsilon_c := sign(\Psi_c)\left(\omega^2 + G_c(B, \mathbb{X}_0)^2\right)^{-1/2}[\omega, G_c(B, \mathbb{X}_0)]'$ is a random vector independent of $w \overset{d}{=} N(0, I_2)$ by Lemma 1 and Assumption 2(i) since $\upsilon_c / \|\upsilon_c\|$ has unit length, (S.19) implies

$$(\upsilon_c / \|\upsilon_c\|)' w \overset{d}{=} N(0,1)$$

as required.

For $\tilde{T}_n(\varphi_n)$ in (S.13), Lemma 2 with $\tilde{z}_0(n) = 0$ implies that

$$\begin{aligned} \tilde{T}_n(\varphi_n) &= \frac{\left|\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1}^0\right|}{\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1}^0}\hat{\sigma}_{\varepsilon}^{-1}\left(\sum_{t=1}^n \tilde{z}_{0,t-1}^2\right)^{-1/2}\sum_{t=1}^n \tilde{z}_{0,t-1}\varepsilon_t + o_p(1) \\ &= sign\left(\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1}^0\right)\frac{1}{\omega}\frac{1}{\sigma_{\varepsilon}}U_n(1) + o_p(1). \end{aligned}$$

By (S.16), $sign\left(\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1}^0\right) \xrightarrow{d} sign(\Psi_c^0)$ where Ψ_c^0 is the restriction of Ψ_c in (S.16) when $G_c(B, \mathbb{X}_0) = 0$. Lemma 1 then implies that $\tilde{T}_n(\varphi_n) \xrightarrow{d} N(0,1)$ as required. This completes the proof of $T_n^{\dagger}(\varphi_n) \xrightarrow{d} N(0,1)$ and $T_n(\varphi_n) \xrightarrow{d} N(0,1)$.

Joint convergence of $\left[T_n^{\dagger}(\varphi_n), T_n(\varphi_n)\right]$ has been established in (S.14) in the weakly persistent case. In particular, (S.14) implies that

$$\max\left\{T_n^{\dagger}(\varphi_n), T_n(\varphi_n)\right\} \xrightarrow{d} \zeta \overset{d}{=} N(0,1).$$

In the strongly persistent case, we may write

$$\begin{bmatrix} T_n^{\dagger}(\varphi_n) \\ T_n(\varphi_n) \end{bmatrix} = \begin{bmatrix} \tilde{T}_n^{\dagger}(\varphi_n) \\ \tilde{T}_n(\varphi_n) \end{bmatrix} + o_p(1) = \mathbb{M}_n \begin{bmatrix} \frac{1}{\omega}\frac{1}{\sigma_{\varepsilon}}U_n(1) \\ \frac{1}{\sigma_{\varepsilon}}\zeta_n(1) \end{bmatrix}. \tag{S.20}$$

where, denoting $\varsigma_n = sign\left\{\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1}\right\}$, $\varsigma_n^0 = sign\left\{\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1}^0\right\}$,

$$\mathbb{M}_n = \begin{bmatrix} \varsigma_n \frac{\omega}{\left(\omega^2 + G_c\left(B_n, n^{-1/2}X_0(n)\right)^2\right)^{1/2}} & \varsigma_n \frac{G_c\left(B_n, n^{-1/2}X_0(n)\right)}{\left(\omega^2 + G_c\left(B_n, n^{-1/2}X_0(n)\right)^2\right)^{1/2}} \\ \varsigma_n^0 & 0 \end{bmatrix}.$$

In the local to unity case, $G_c\left(B_n, n^{-1/2}X_0\left(n\right)\right) \overset{d}{\to} G_c\left(B, \mathbb{X}_0\right) \neq 0$ $a.s.$ since $G_c\left(B, \mathbb{X}_0\right)$ is continuously distributed by Assumption 2. Also, $\varsigma_n \overset{d}{\to} \varsigma_c := sign\left\{\Psi_c\right\}$, $\varsigma_n^0 \overset{d}{\to} \varsigma_c^0 := sign\left\{\Psi_c^0\right\}$. By applying Lemma 1 to (S.20), $\left[\frac{1}{\omega}\frac{1}{\sigma_\varepsilon}U_n\left(1\right), \frac{1}{\sigma_\varepsilon}\zeta_n\left(1\right)\right]' \overset{d}{\to} N\left(0, I_2\right)$, $\mathbb{M}_n \overset{d}{\to} \mathbb{M}$ and

$$\begin{bmatrix} T_n^{\dagger}\left(\varphi_n\right) \\ T_n\left(\varphi_n\right) \end{bmatrix} \overset{d}{\underline{=}} \begin{bmatrix} \frac{\omega}{\left(\omega^2+G_c^2\right)^{1/2}}Z_1 + \frac{G_c}{\left(\omega^2+G_c^2\right)^{1/2}}Z_2 \\ Z_1 \end{bmatrix} \tag{S.21}$$

where $Z_1 = \frac{\varsigma_c}{\omega\sigma_\varepsilon}U\left(1\right)$ and $Z_2 = \frac{\varsigma_c^0}{\sigma_\varepsilon}\zeta\left(1\right)$ are independent $N\left(0,1\right)$ random variables, by independence of ς_c and $U\left(1\right)$ and of ς_c^0 and $\zeta\left(1\right)$ and we abbreviate $G_c := G_c\left(B, \mathbb{X}_0\right)$. Next, we obtain an expression for the distribution of

$$T_{\max} := \max\left\{\frac{\omega}{\left(\omega^2+G_c^2\right)^{1/2}}Z_1 + \frac{G_c}{\left(\omega^2+G_c^2\right)^{1/2}}Z_2, Z_1\right\},$$

the maximum of the limits in distribution of $\left[T_n^{\dagger}\left(\varphi_n\right), T_n\left(\varphi_n\right)\right]$ in (S.21). The c.d.f. of $T_{\max}$ is given by

$$\begin{aligned} \mathbb{P}\left(T_{\max} \leq \lambda\right) &= \mathbb{P}\left(Z_1 \leq \frac{\omega}{\left(\omega^2+G_c^2\right)^{1/2}}Z_1 + \frac{G_c}{\left(\omega^2+G_c^2\right)^{1/2}}Z_2 \leq \lambda\right) \\ &\quad + \mathbb{P}\left(\frac{\omega}{\left(\omega^2+G_c^2\right)^{1/2}}Z_1 + \frac{G_c}{\left(\omega^2+G_c^2\right)^{1/2}}Z_2 \leq Z_1 \leq \lambda\right) \\ &= p_1\left(\lambda\right) + p_2\left(\lambda\right) \end{aligned}$$

where

$$p_1\left(\lambda\right) = \mathbb{P}\left(\left[\left(\omega^2+G_c^2\right)^{1/2} - \omega\right]Z_1 \leq G_cZ_2 \leq \left(\omega^2+G_c^2\right)^{1/2}\lambda - \omega Z_1\right)$$

and

$$p_2\left(\lambda\right) = \mathbb{P}\left(G_cZ_2 \leq \left(\left(\omega^2+G_c^2\right)^{1/2} - \omega\right)Z_1, Z_1 \leq \lambda\right).$$

Denoting $A_c := \left(\omega^2+G_c^2\right)^{1/2} - \omega$, $B_c := \left(\omega^2+G_c^2\right)^{1/2}$ and using the law of iterated expectations, we obtain

$$\begin{aligned} \mathbb{P}\left(T_{\max} \leq \lambda\right) &= \mathbb{P}\left(A_cZ_1 \leq G_cZ_2 \leq B_c\lambda - \omega Z_1\right) + \mathbb{P}\left(G_cZ_2 \leq A_cZ_1, Z_1 \leq \lambda\right) \\ &= \mathbb{E}\left[\mathbb{P}\left(A_cZ_1 \leq G_cZ_2 \leq B_c\lambda - \omega Z_1 \middle| G_c\right)\right] \\ &\quad + \mathbb{E}\left[\mathbb{P}\left(G_cZ_2 \leq A_cZ_1, Z_1 \leq \lambda \middle| G_c\right)\right]. \end{aligned} \tag{S.22}$$

Since Z_1 is independent of G_cZ_2, the density of (Z_1, G_cZ_2) conditional on G_c is given by

$$f_{(Z_1,G_cZ_2)}\left(x_1,x_2\right) = \frac{1}{2\pi}e^{-\frac{1}{2}x_1^2}e^{-\frac{1}{2G_c^2}x_2^2}.$$

Hence,

$$\begin{aligned}\mathbb{P}\left(A_cZ_1 \le G_cZ_2 \le B_c\lambda - \omega Z_1 \middle| G_c\right) &= \int_{-\infty}^{\infty}\int_{A_cx_1}^{B_c\lambda-\omega x_1} f_{(Z_1,G_cZ_2)}\left(x_1,x_2\right)dx_2dx_1 \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-\frac{1}{2}x_1^2}\int_{A_cx_1}^{B_c\lambda-\omega x_1}e^{-\frac{1}{2G_c^2}x_2^2}dx_2dx_1\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}\left(G_cZ_2 \le A_cZ_1, Z_1 \le \lambda \middle| G_c\right) &= \int_{-\infty}^{\lambda}\int_{-\infty}^{A_cx_1} f_{(Z_1,G_cZ_2)}\left(x_1,x_2\right)dx_2dx_1 \\ &= \frac{1}{2\pi}\int_{-\infty}^{\lambda}e^{-\frac{1}{2}x_1^2}\int_{-\infty}^{A_cx_1}e^{-\frac{1}{2G_c^2}x_2^2}dx_2dx_1\end{aligned}$$

Substituting into (S.22) and differentiating with respect to λ, we obtain the density function of $T_{\max}$:

$$\begin{aligned}f_{T_{\max}}\left(\lambda\right) &= \frac{1}{2\pi}\mathbb{E}\left(B_c\int_{-\infty}^{\infty}e^{-\frac{1}{2}x_1^2}e^{-\frac{1}{2G_c^2}\left(B_c\lambda-\omega x_1\right)^2}dx_1 + e^{-\frac{1}{2}\lambda^2}\int_{-\infty}^{A_c\lambda}e^{-\frac{1}{2G_c^2}x_2^2}dx_2\right) \\ &= \frac{1}{2\pi}\mathbb{E}B_ce^{-\frac{1}{2G_c^2}B_c^2\lambda^2}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\left(1+\frac{\omega^2}{G_c^2}\right)\left[x_1^2-2\frac{B_c\lambda\omega}{G_c^2+\omega^2}x_1\right]}dx_1 \\ &\quad+\frac{1}{2\pi}e^{-\frac{1}{2}\lambda^2}\mathbb{E}\int_{-\infty}^{A_c\lambda}e^{-\frac{1}{2}\left(\frac{x_2}{|G_c|}\right)^2}dx_2 \\ &= \frac{1}{2\pi}\mathbb{E}B_ce^{-\frac{1}{2G_c^2}B_c^2\lambda^2}e^{\frac{1}{2}\left(1+\frac{\omega^2}{G_c^2}\right)\left(\frac{B_c\lambda\omega}{G_c^2+\omega^2}\right)^2}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\left(1+\frac{\omega^2}{G_c^2}\right)\left(x_1-\frac{B_c\lambda\omega}{G_c^2+\omega^2}\right)^2}dx_1 \\ &\quad+\frac{1}{2\pi}e^{-\frac{1}{2}\lambda^2}\mathbb{E}\left|G_c\right|\int_{-\infty}^{\frac{A_c\lambda}{|G_c|}}e^{-\frac{1}{2}x_2^2}dx_2 \\ &= \frac{1}{\sqrt{2\pi}}\mathbb{E}\frac{B_ce^{-\frac{1}{2}\frac{B_c^2\lambda^2}{G_c^2+\omega^2}}}{\sqrt{1+\omega^2/G_c^2}} + \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\lambda^2}\mathbb{E}\left[\left|G_c\right|\Phi\left(\frac{A_c\lambda}{|G_c|}\right)\right] \\ &= \phi\left(\lambda\right)\left[\mathbb{E}\left|G_c\right| + \mathbb{E}\left|G_c\right|\Phi\left(\lambda\frac{\left(\omega^2+G_c^2\right)^{1/2}-\omega}{|G_c|}\right)\right],\end{aligned}$$

as required.

Additional Reference

Jacod, J. and Shiryaev, A.N.S. (2003). Limit theorems for stochastic processes. Springer-Verlag Berlin Heidelberg New York.

S.2 Residual Wild Bootstrap Implementation

We apply the residual wild bootstrap [RWB] method detailed in Algorithm 1 of Demetrescu *et al.* (2022, p.13) to obtain simulated bootstrap critical values for the $T_n(\varphi_n)$, $T_n^{\dagger}(\varphi_n)$, $T_n(\varphi_n)^2$ and $T_n^{\dagger}(\varphi_n)^2$ tests considered in this paper. In Algorithm S.1, we outline the procedure to implement the $T_n^{\dagger}(\varphi_n)$ test statistic at the α significance level. The corresponding RWB implementation of the other tests is performed in the same manner; see the subsequent discussion in Remarks S.2.1–S.2.3. Details on how to obtain simulated bootstrap p-values for the tests are given in Remark S.2.4. Then in section S.2.1 we detail how to obtain simulated bootstrap critical values and p-values for the union-of-rejection tests discussed in Remark 3.3.

Algorithm S.1. ***(Residual Wild Bootstrap Algorithm)***

Step 1: *Estimate (1) by OLS to obtain the residuals $\hat{\varepsilon}_t$, $t = 1, ..., n$.*

Step 2: *Estimate by OLS the equation*

$$x_t = m + \sum_{j=1}^{p} a_j x_{t-j} + b\tilde{y}_t + u_t \tag{S.23}$$

where $\tilde{y}_t := y_t - n^{-1}\sum_{j=1}^{n} y_j$, to obtain the OLS estimates $\hat{m}$, $\hat{a}_j$, $j = 1, ..., p$, and $\hat{b}$, and compute $\hat{u}_t := x_t - \hat{m} - \sum_{j=1}^{p} \hat{a}_j x_{t-j}$, $t = p, ..., n$. Set $\hat{u}_t = 0$ for $t = 1, ..., p-1$.

Step 3: *Generate bootstrap innovations $(\varepsilon_t^*, u_t^*)' := (R_t\hat{\varepsilon}_t, R_t\hat{u}_t)'$, $t = 1, ..., n$ where R_t $t = 1, ..., n$ is a scalar sequence of $NIID(0,1)$ random variables which are independent of the sample data.*

Step 4: *Define the bootstrap data $(y_t^*, x_{t-1}^*)'$, $t = 1, ..., n$, where $y_t^* = \varepsilon_t^*$ (so that the null hypothesis is imposed on the bootstrap data y_t^*) and where x_t^* is generated according to the recursion*

$$x_t^* = \sum_{j=1}^{p} \hat{a}_j x_{t-j}^* + u_t^*, \quad t = 1, ..., n$$

with initial conditions $x_0^* = ... = x_{-(p-1)}^* = 0$. *Create the associated IVX instrument,* $\tilde{z}_t^* = \varphi_n \tilde{z}_{t-1}^* + \Delta x_t^*$, $t = 1, ..., n$, *initialised at* $\tilde{z}_0^* = x_0^* - \bar{x}_n^*$, *where* $\bar{x}_n^* = n^{-1} \sum_{t=1}^n x_t^*$. *Notice that* φ_n *is the same value used in constructing the IVX instrument,* $\tilde{z}_t$, *from* x_t.

Step 5: *Using the bootstrap data* $(y_t^*, x_{t-1}^*, \tilde{z}_{t-1}^*)'$, $t = 1, ..., n$, *in place of the original sample data,* $(y_t, x_{t-1}, \tilde{z}_{t-1})'$, $t = 1, ..., n$, *construct the bootstrap analogue of the* $T_n^{\dagger}(\varphi_n)$ *statistic, denoting this test statistic as* $T_n^{\dagger}(\varphi_n)^*$.

Step 6: *Repeat steps 3-5 B times, with the* $\{R_t\}$ *sequences additionally independent across the B bootstrap replications, and define* $cv_{T_n(\varphi_n)^*,\alpha}$ *and* $cv_{T_n(\varphi_n)^*,1-\alpha}$ *as the* α *and* $(1-\alpha)$ *quantiles, respectively, of the resulting B bootstrap test statistics. The null is rejected at the* α *significance level when testing in the right tail if the* $T_n(\varphi_n)$ *test statistic is greater than* $cv_{T_n(\varphi_n)^*,1-\alpha}$, *and when testing in the left tail the null is rejected at the* α *significance level if the* $T_n(\varphi_n)$ *test statistic is less than* $cv_{T_n(\varphi_n)^*,\alpha}$.

Remark S.2.1. Notice that Algorithm S.1 differs slightly from Algorithm 1 of Demetrescu *et al.* (2022) in that $\tilde{y}_t$ is included as a covariate in (S.23) in Step 2. This results in improved estimates of the autoregressive parameters from this regression, but does not alter the large sample properties of the bootstrap statistics relative to Algorithm 1 of Demetrescu *et al.* (2022) where $\tilde{y}_t$ is not included in (S.23). ◇

Remark S.2.2. A bootstrap implementation of $T_n(\varphi_n)$ can be performed in a similar manner to that detailed for $T_n^{\dagger}(\varphi_n)$ in Algorithm S.1. The only difference would be that the bootstrap instrument $\tilde{z}_t^*$ would be initialised at $\tilde{z}_0^* = 0$. For future reference, we define $cv_{T_n^{\dagger}(\varphi_n)^*,\alpha}$ and $cv_{T_n^{\dagger}(\varphi_n)^*,1-\alpha}$ as the α and $(1-\alpha)$ quantiles of the resulting B bootstrap test statistics. ◇

Remark S.2.3. A two-sided variant of both tests can be performed by replacing $T_n^{\dagger}(\varphi_n)$ or $T_n(\varphi_n)$, and their bootstrap analogues, with $T_n^{\dagger}(\varphi_n)^2$ or $T_n(\varphi_n)^2$, respectively, and rejecting in the upper tail. ◇

Remark S.2.4. We can also construct simulated bootstrap p-values for the $T_n(\varphi_n)$ and $T_n^{\dagger}(\varphi_n)$ tests. Taking the $T_n(\varphi_n)$ statistic to illustrate the method, denote the B bootstrap analogue statistics obtained from running Algorithm S.1 as $T_{n,i}(\varphi_n)^*$, $i = 1, ..., B$. The simulated upper tail bootstrap p-value for the test is then calculated as the percentage of bootstrap statistics that do not exceed the actual statistic; that is, $p^U_{T_n(\varphi_n)} := B^{-1}\sum_{i=1}^{B}\mathbb{I}(T_{n,i}(\varphi_n)^* < T_n(\varphi_n))$. Similarly, the simulated lower tail bootstrap p-value is calculated as the percentage of bootstrap statistics that exceed the actual statistic; that is, $p^L_{T_n(\varphi_n)} := B^{-1}\sum_{i=1}^{B}\mathbb{I}(T_{n,i}(\varphi_n)^* > T_n(\varphi_n))$. As $B \to \infty$, these both converge almost surely to the true p-values. ♢

Remark S.2.5. We also considered initialising the bootstrap series x_t^* at $x_0^* = x_0 - \bar{x}_n$ but found that while this did improve the size of tests based on $T_n(\varphi_n)$ when $c, \alpha > 0$, they were still significantly less powerful than tests based on $T_n^\dagger(\varphi_n)$. ♢

S.2.1 Union of Rejection Tests

The methods outlined above are applicable for tests based on $T_n(\varphi_n)$, $T_n^\dagger(\varphi_n)$, $T_n(\varphi_n)^2$ and $T_n^\dagger(\varphi_n)^2$. Simulated bootstrap critical values and p-values for the union-of-rejections tests discussed in Remark 3.3 are simple to obtain using the approach outlined in section 2.3 of Smeekes and Taylor (2012) for bootstrapping union of rejections type tests. Usefully, these not require any additional bootstrap computation.

(i) Consider first the right-tailed test for $H_0 : \beta = 0$ against $H_1 : \beta > 0$ based on the maximum of the $T_n(\varphi_n)$ and $T_n^\dagger(\varphi_n)$ statistics, which rejects for large values of the U_R statistic defined in Remark 3.3. Our bootstrap implementation of this test is based on the statistic,

$$\underline{U}_R := \max\left\{T_n^\dagger(\varphi_n), T_n(\varphi_n) + d_R^*\right\}$$

where $d_R^* := cv_{T_n^\dagger(\varphi_n)^*,1-\alpha} - cv_{T_n(\varphi_n)^*,1-\alpha}$, in which $cv_{T_n^\dagger(\varphi_n)^*,1-\alpha}$ is the α level upper-tailed bootstrap critical value for $T_n^\dagger(\varphi_n)$ defined in Step 6 of Algorithm S.1, and $cv_{T_n(\varphi_n)^*,1-\alpha}$ is the corresponding quantity for $T_n(\varphi_n)$ defined in Remark S.2.2. The correction term, d_R^*, is added onto $T_n(\varphi_n)$ to allow for the possibility that the finite sample null distributions of $T_n^\dagger(\varphi_n)$ and $T_n(\varphi_n)$ are not equal at the $1-\alpha$ quantile. In large samples d_R^* will be zero because the two statistics have the same limiting null distribution. The use of an additive correction term is a slight modification to the union-of-rejections strategy proposed by Smeekes and Taylor (2012) who use a multiplicative correction factor to line up the finite sample critical values of the statistics forming the union they consider.[1]

Denoting the ith bootstrap analogues of the $T_n(\varphi_n)$, $T_n^{\dagger}(\varphi_n)$ statistics, obtained as in Algorithm S.1, as $T_{n,i}(\varphi_n)^*$ and $T_{n,i}^{\dagger}(\varphi_n)^*$, $i = 1, ..., B$, one can compute the corresponding ith bootstrap statistic, $\underline{U}_{R,i}^*$, as

$$\underline{U}_{R,i}^* := \max\left\{T_{n,i}^{\dagger}(\varphi_n)^*, T_{n,i}(\varphi_n)^* + d_R^*\right\}.$$

The $(1-\alpha)$ quantile of the B bootstrap $\underline{U}_{R,i}^*$, $i = 1, ..., B$, statistics then gives a simulated bootstrap critical value appropriate for right-tailed testing at the α level based on the U_R test statistic.

(ii) Next consider the left-tailed test for $H_0 : \beta = 0$ against $H_1 : \beta < 0$ based on the minimum of the $T_n(\varphi_n)$ and $T_n^{\dagger}(\varphi_n)$ statistics which rejects for large negative values of the U_L statistic in Remark 3.3. A bootstrap implementation of this test is based on the statistic

$$\underline{U}_L := \min\left\{T_n^{\dagger}(\varphi_n), T_n(\varphi_n) + d_L^*\right\}$$

where $d_L^* := cv_{T_n^{\dagger}(\varphi_n)^*,\alpha} - cv_{T_n(\varphi_n)^*,\alpha}$.

Again, denoting the ith bootstrap analogues of the $T_n(\varphi_n)$, $T_n^{\dagger}(\varphi_n)$ statistics, obtained as in Algorithm S.1, as $T_{n,i}(\varphi_n)^*$ and $T_{n,i}^{\dagger}(\varphi_n)^*$, $i = 1, ..., B$, one can compute the corresponding ith bootstrap statistic, $\underline{U}_{L,i}^*$ as

$$\underline{U}_{L,i}^* := \min\left\{T_{n,i}^{\dagger}(\varphi_n)^*, T_{n,i}(\varphi_n)^* + d_L^*\right\}$$

[1]In the context of left-tailed tests we found that the two bootstrap critical values used to calculate d_R^* can sometimes have opposite signs, necessitating the additive, rather than multiplicative, correction term.

The α quantile of the B bootstrap $\underline{U}^*_{L,i}$, $i = 1, ..., B$, statistics then gives a simulated bootstrap critical value appropriate for left-tailed testing at the α level based on the U_L test statistic.

(iii) Finally, consider the two-sided test for $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ based on the maximum of the $T_n(\varphi_n)^2$ and $T_n^\dagger(\varphi_n)^2$ statistics which rejects for large positive values of the U_{2S} statistic in Remark 3.3. Denoting the α level upper-tailed bootstrap critical values for $T_n(\varphi_n)^2$ and $T_n^\dagger(\varphi_n)^2$ as $cv_{T_n(\varphi_n)^{*2},1-\alpha}$ and $cv_{T_n^\dagger(\varphi_n)^{*2},1-\alpha}$, respectively, a bootstrap implementation of the test can be based on the statistic

$$\underline{U}_{2S} := \max\left\{T_n^\dagger(\varphi_n)^2, T_n(\varphi_n)^2 + d^*_{2S}\right\}$$

where $d^*_{2S} := cv_{T_n^\dagger(\varphi_n)^{*2},1-\alpha} - cv_{T_n(\varphi_n)^{*2},1-\alpha}$.

Denoting the ith bootstrap analogues of the $T_n(\varphi_n)^2$, $T_n^\dagger(\varphi_n)^2$ statistics, obtained as in Algorithm S.1, as $T_{n,i}(\varphi_n)^{*2}$ and $T^\dagger_{n,i}(\varphi_n)^{*2}$, $i = 1, ..., B$, one can compute the corresponding ith bootstrap statistic, $\underline{U}^*_{2S,i}$ as

$$\underline{U}^*_{2S,i} := \min\left\{T^\dagger_{n,i}(\varphi_n)^{*2}, T_{n,i}(\varphi_n)^{*2} + d^*_{2S}\right\}$$

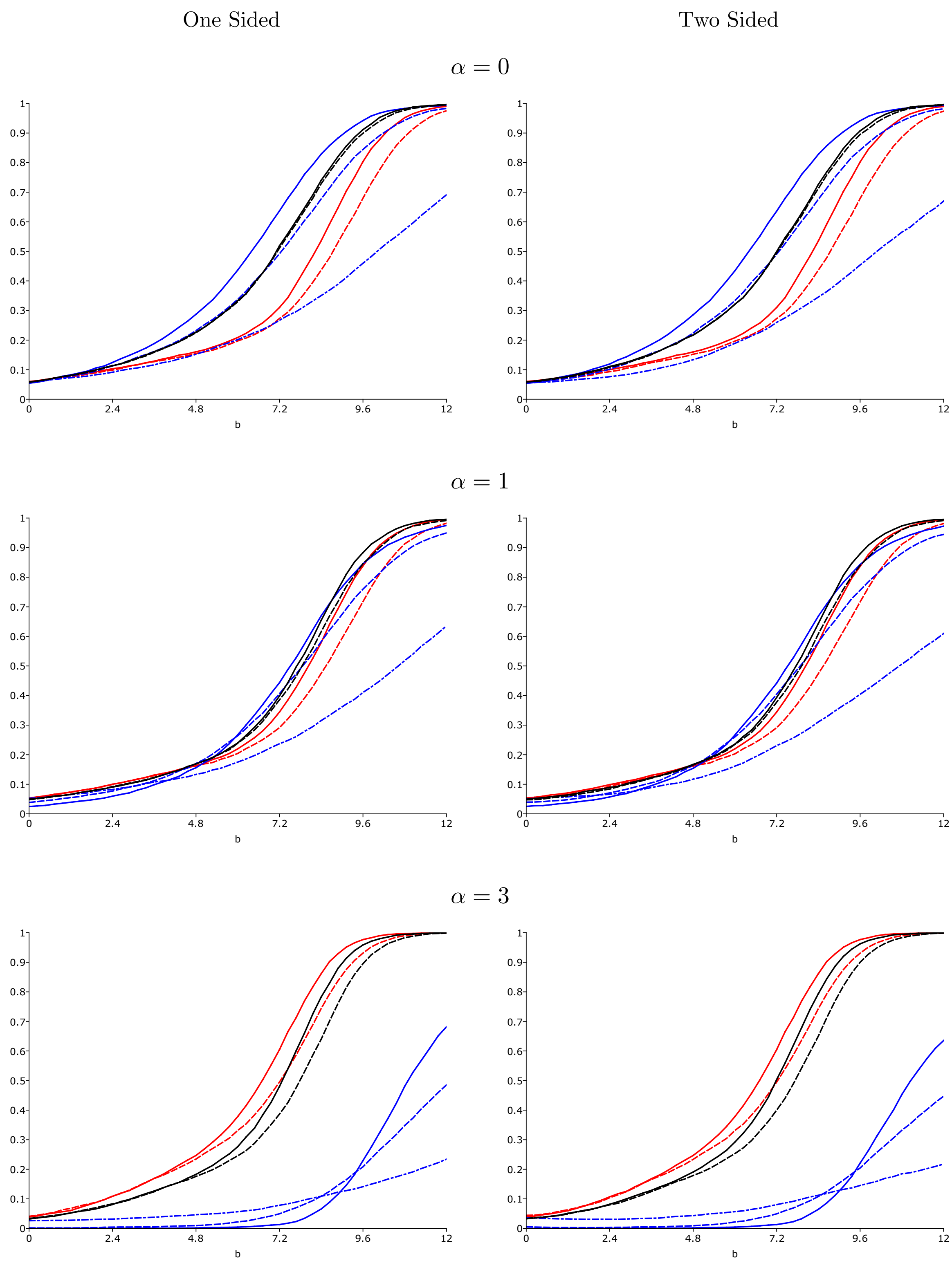
where $d^*_{2S} := cv_{T_n^\dagger(\varphi_n)^{*2},1-\alpha} - cv_{T_n(\varphi_n)^{*2},1-\alpha}$, with $cv_{T_n^\dagger(\varphi_n)^{*2},1-\alpha}$ and $cv_{T_n(\varphi_n)^{*2},1-\alpha}$ the simulated $(1-\alpha)$ bootstrap quantiles obtained from the ordered $T^\dagger_{n,i}(\varphi_n)^{*2}$ and $T_{n,i}(\varphi_n)^{*2}$, $i = 1, ..., B$, bootstrap statistics, respectively. The $(1-\alpha)$ quantile of the B bootstrap $\underline{U}^*_{2S,i}$, $i = 1, ..., B$, statistics then gives a simulated bootstrap critical value appropriate for two-sided testing at the α level based on the U_{2S} test statistic.

For each of the bootstrap test procedures outlined in (i)–(iii) above, corresponding simulated bootstrap p-values can be calculated by performing the bootstrap variant of the test over a fine grid of different significance levels, with the bootstrap p-value chosen as the lowest (closest to zero) significance level for which the test signals a rejection.

S.3 Additional Monte Carlo Simulation Results

In this section we provide additional Monte Carlo results to those provided in Section 5 for: $c \in \{5, 10, 30, 40\}$ for $n = 250$; and $c \in \{0, 2, 5, 10, 20, 30, 40, 50\}$ for $n = 1000$.

Figure S.1: Finite Sample Power, Right Tail, $n = 250$, $\delta = -0.95$, $c = 5$



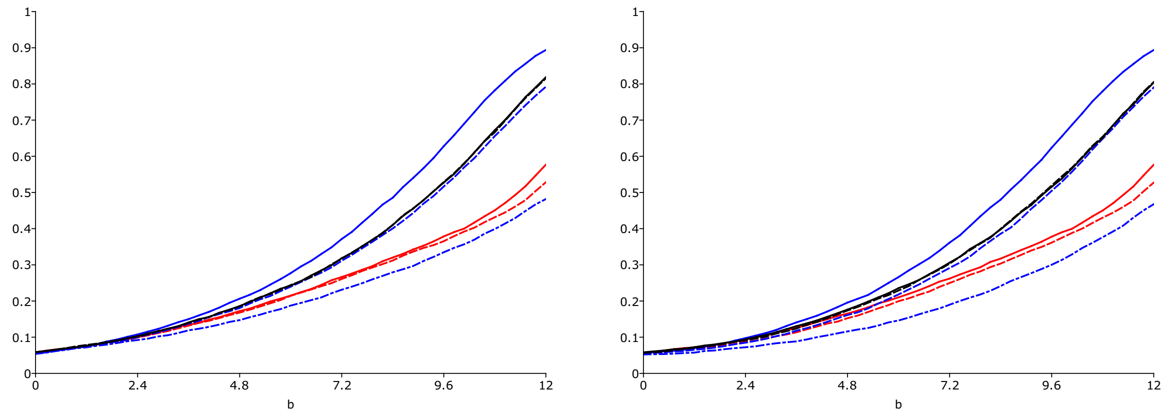
$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

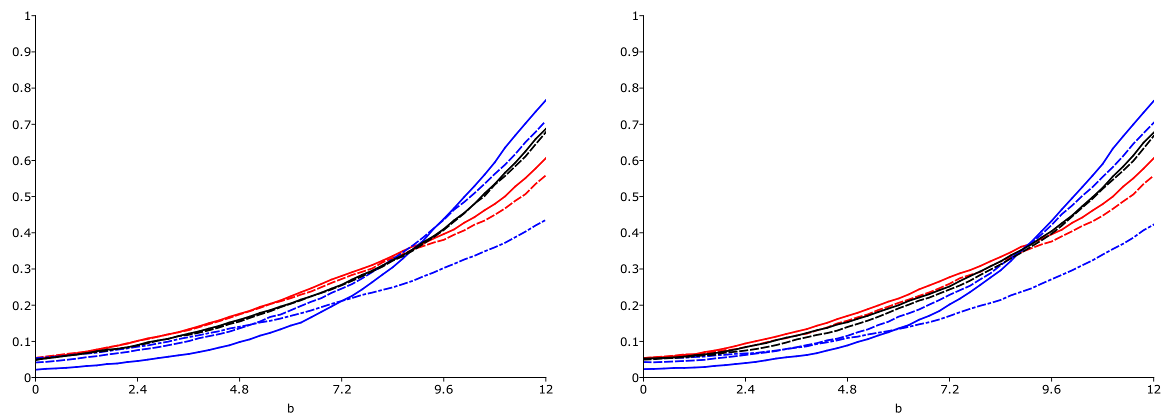
Figure S.2: Finite Sample Power, Right Tail, $n = 250$, $\delta = -0.95$, $c = 10$

One Sided | Two Sided

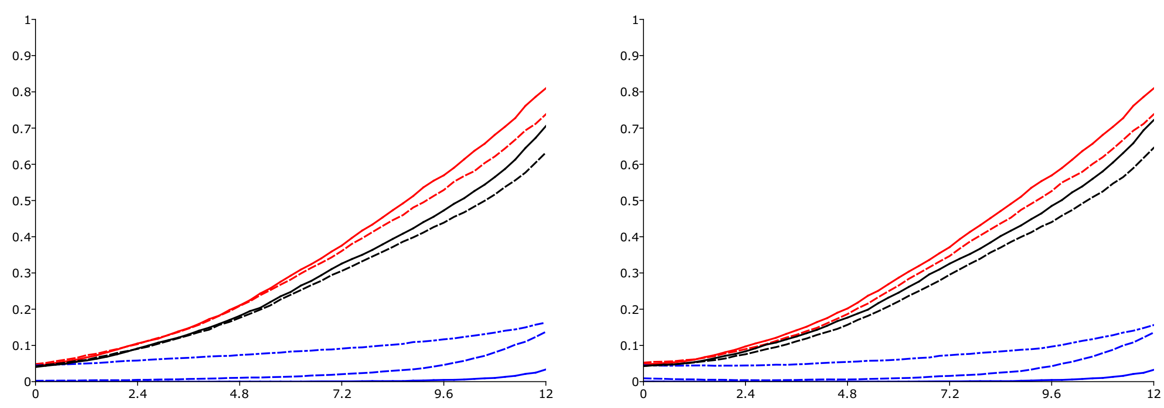
$\alpha = 0$



$\alpha = 1$



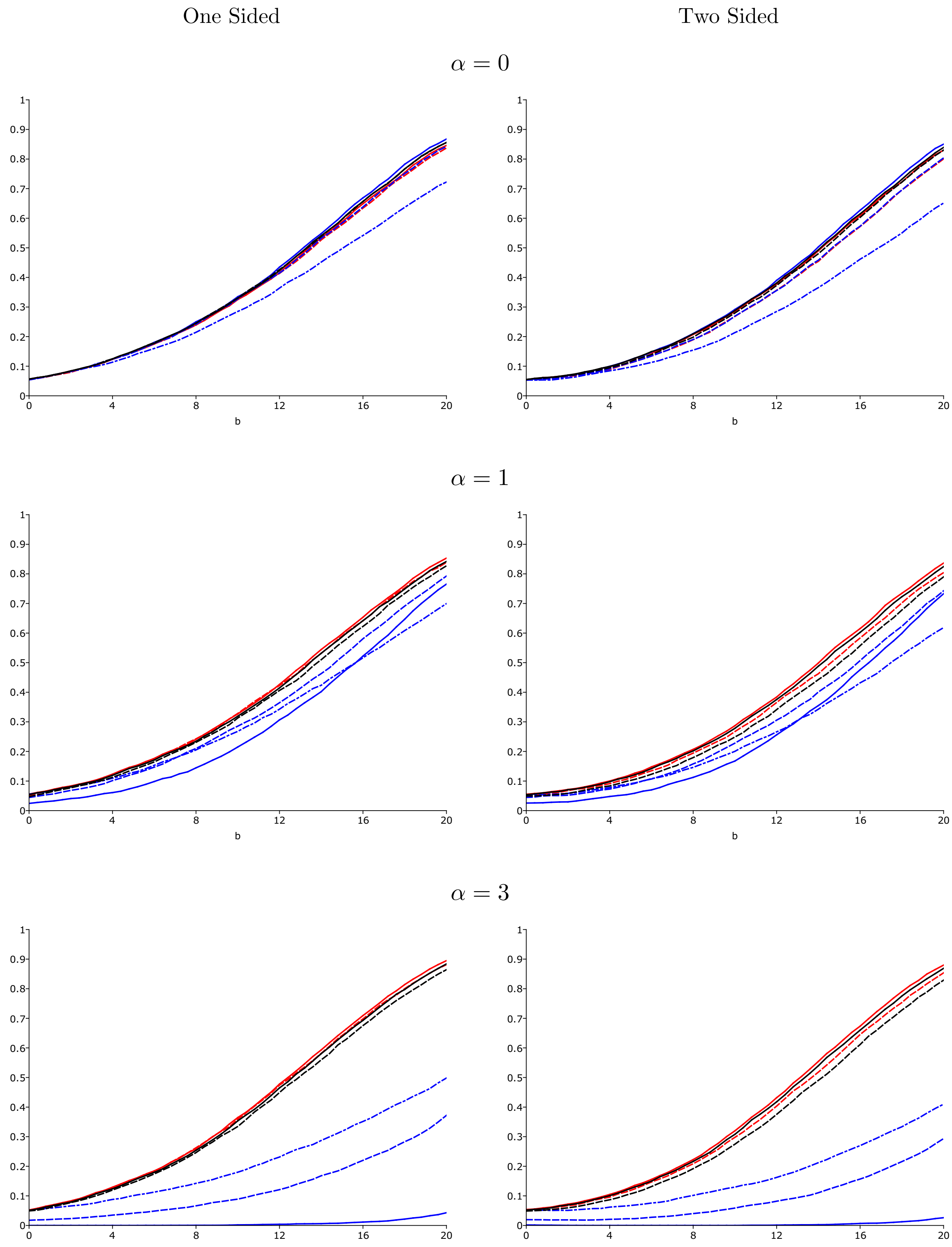
$\alpha = 3$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

Figure S.3: Finite Sample Power, Right Tail, $n = 250$, $\delta = -0.95$, $c = 30$

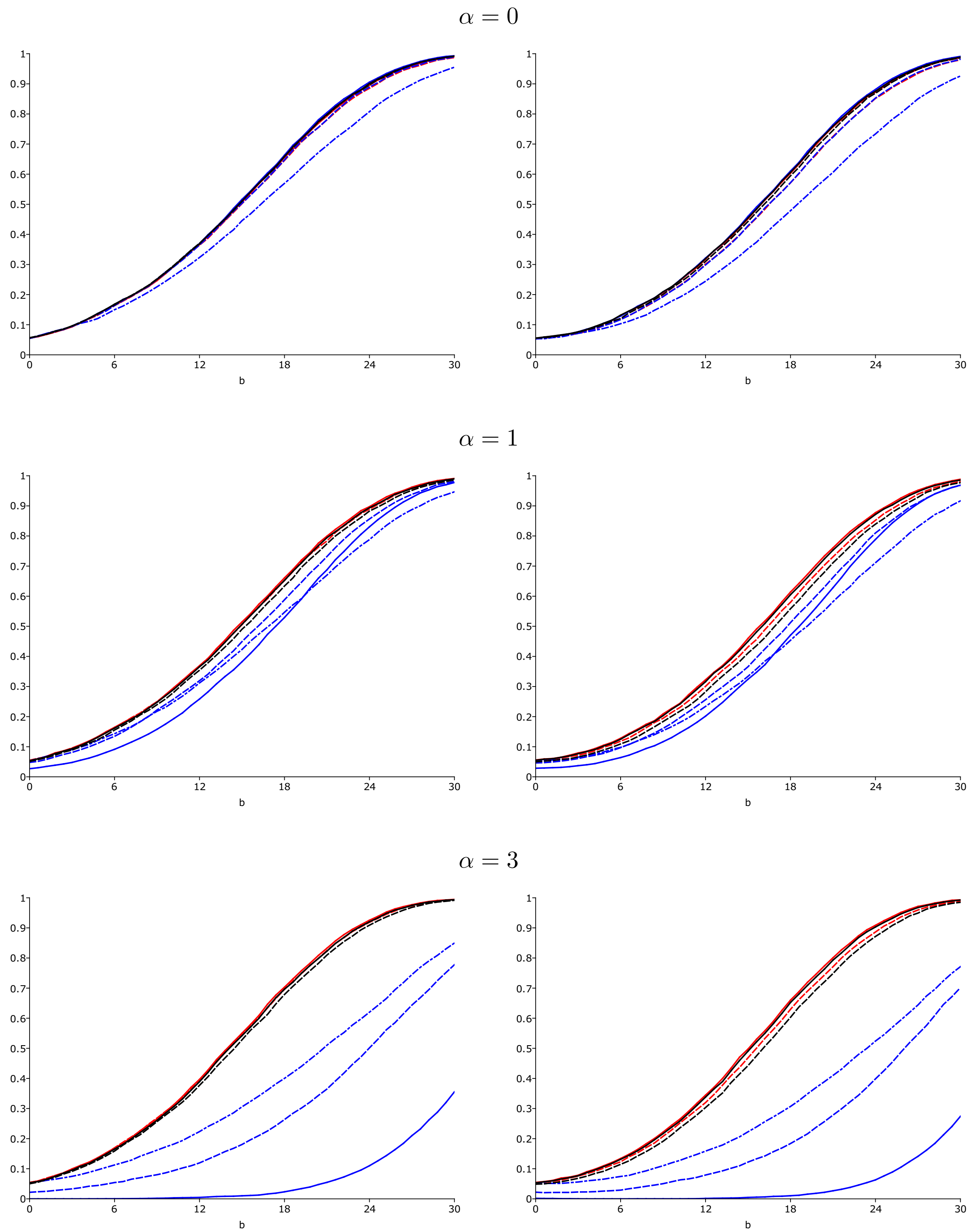


$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

Figure S.4: Finite Sample Power, Right Tail, $n = 250$, $\delta = -0.95$, $c = 40$

One Sided Two Sided



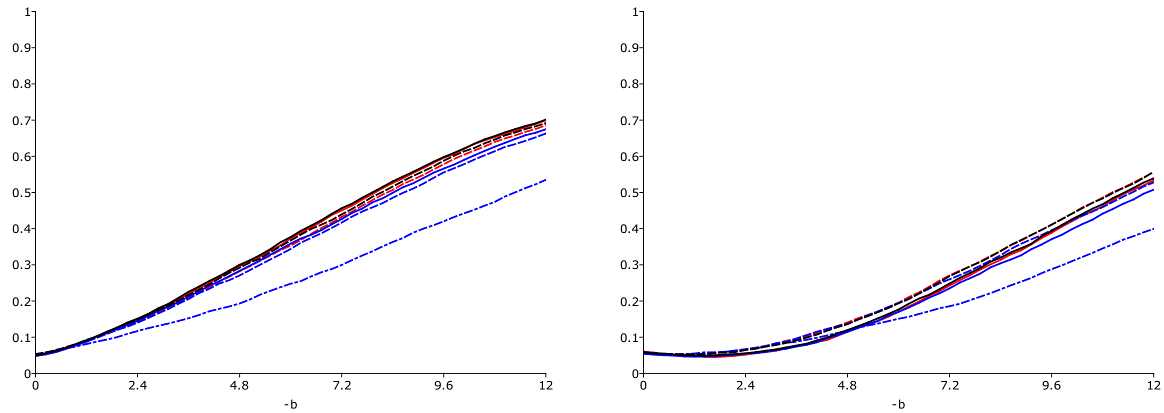
$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

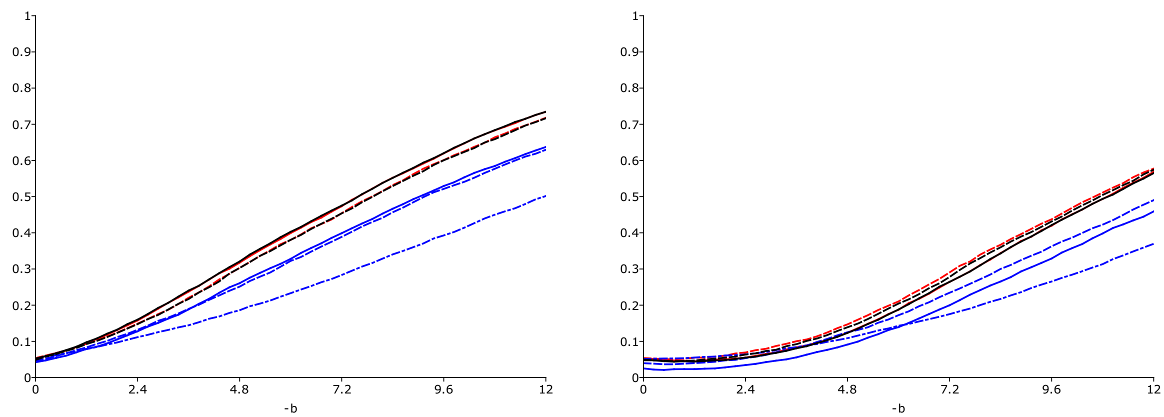
Figure S.5: Finite Sample Power, Left Tail, $n = 250$, $\delta = -0.95$, $c = 5$

One Sided | Two Sided

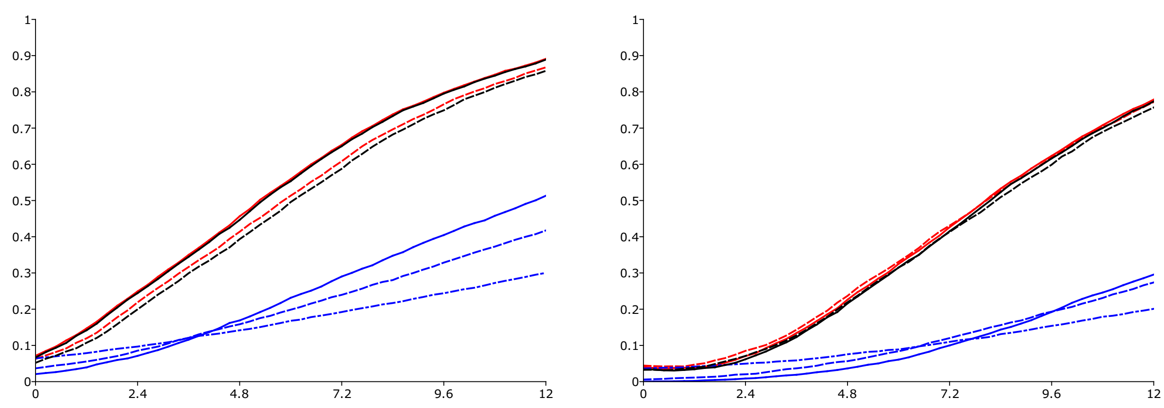
$\alpha = 0$



$\alpha = 1$



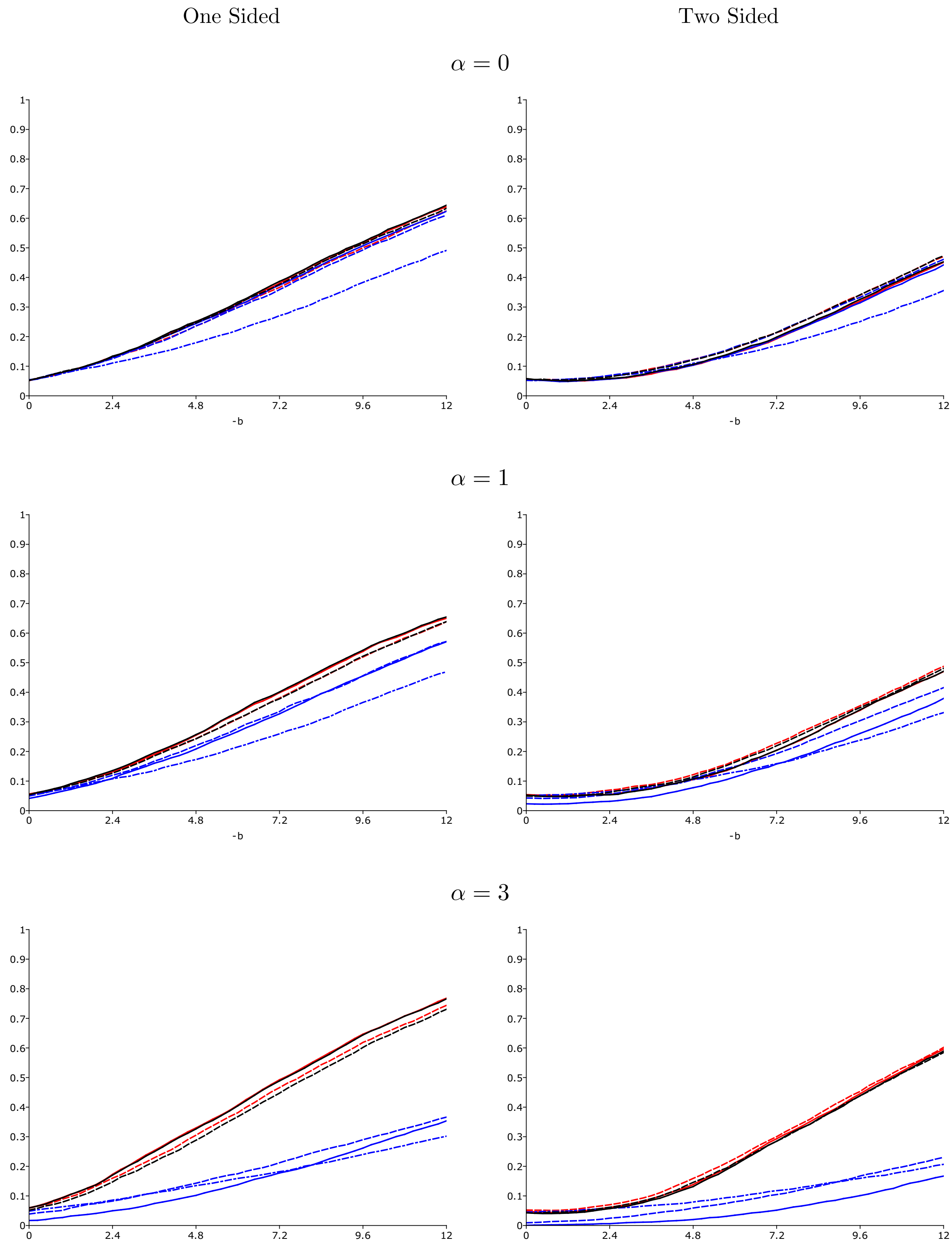
$\alpha = 3$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

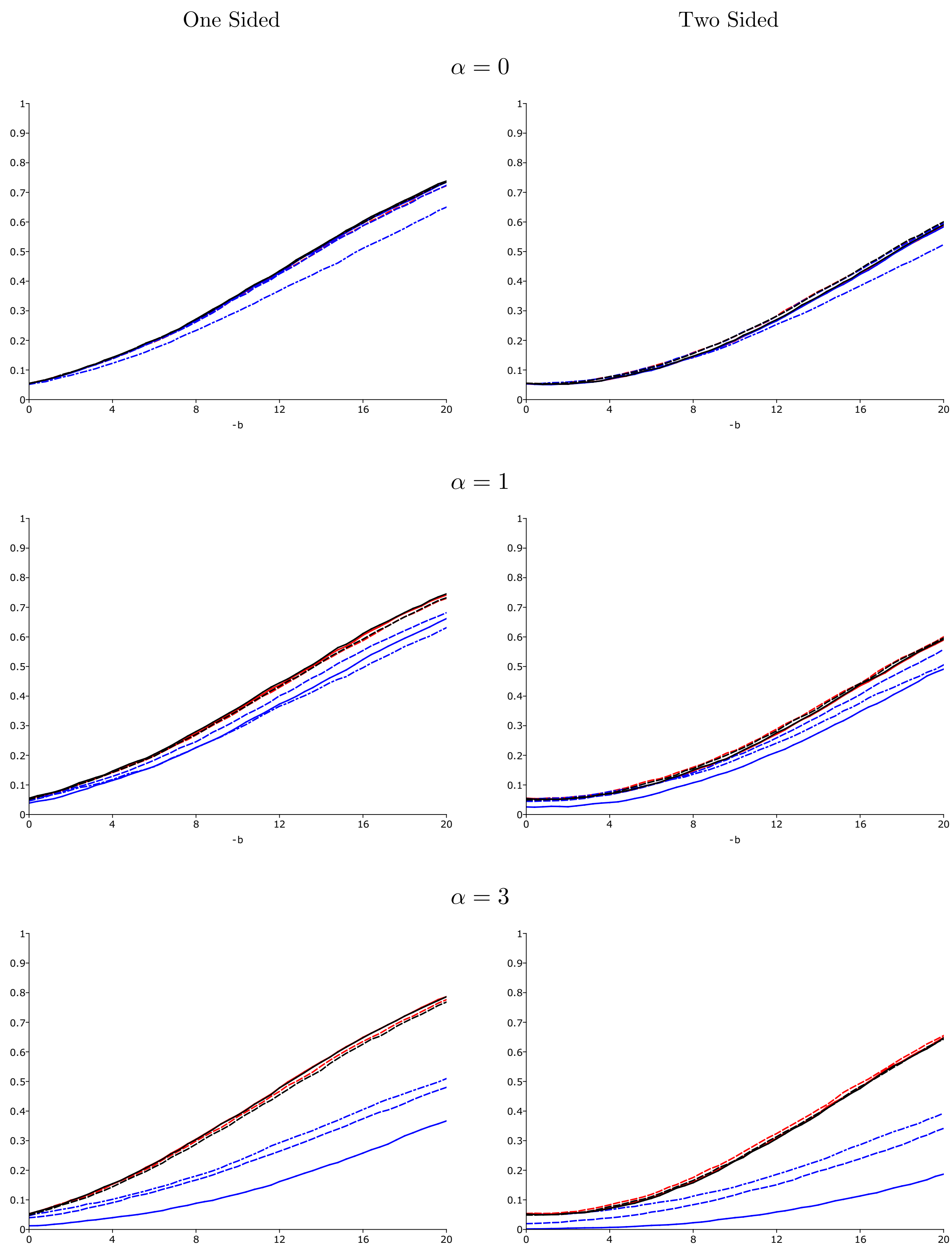
Figure S.6: Finite Sample Power, Left Tail, $n = 250$, $\delta = -0.95$, $c = 10$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

Figure S.7: Finite Sample Power, Left Tail, $n = 250$, $\delta = -0.95$, $c = 30$



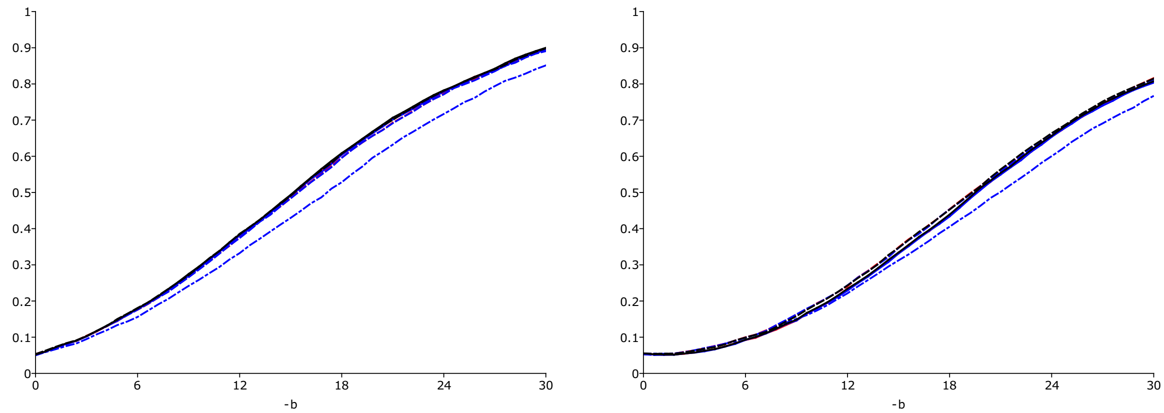
$IVX_{0.95}$: ———, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ———, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ———, $U_{0.75}$: – –

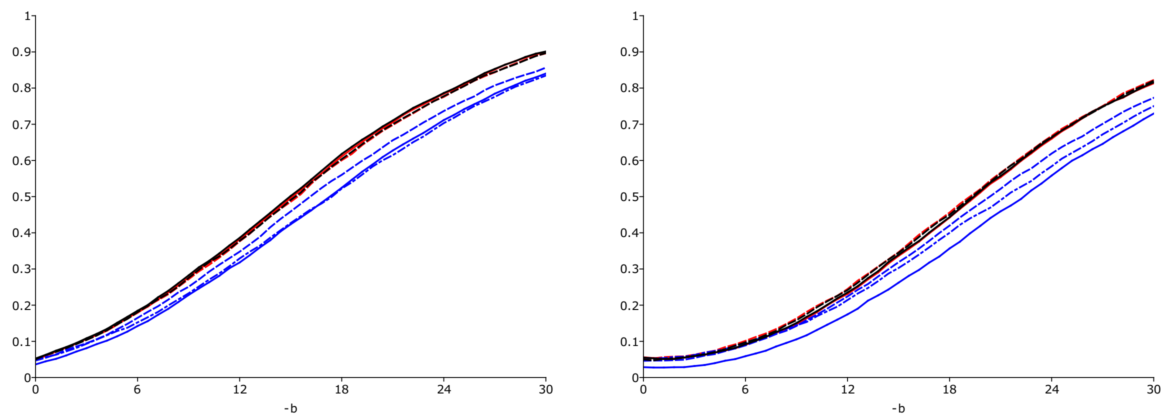
Figure S.8: Finite Sample Power, Left Tail, $n = 250$, $\delta = -0.95$, $c = 40$

One Sided | Two Sided

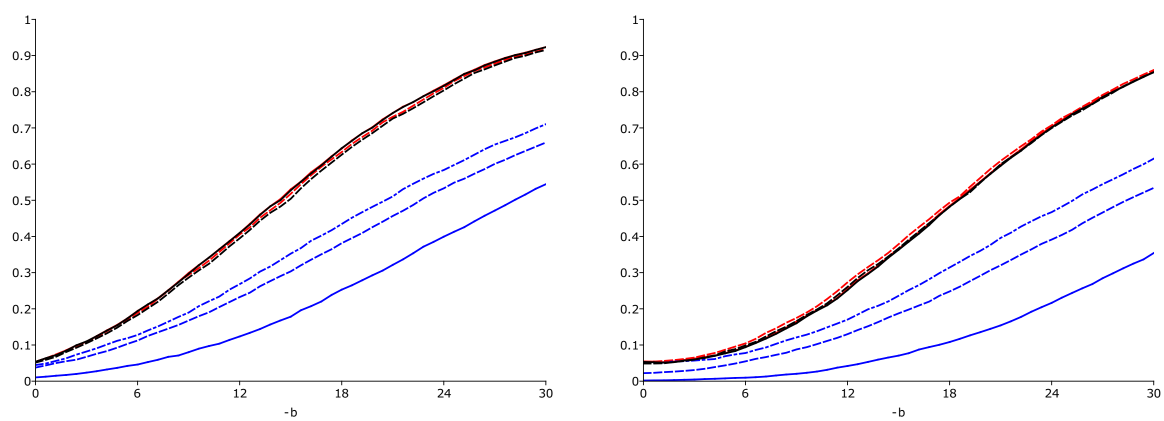
$\alpha = 0$



$\alpha = 1$



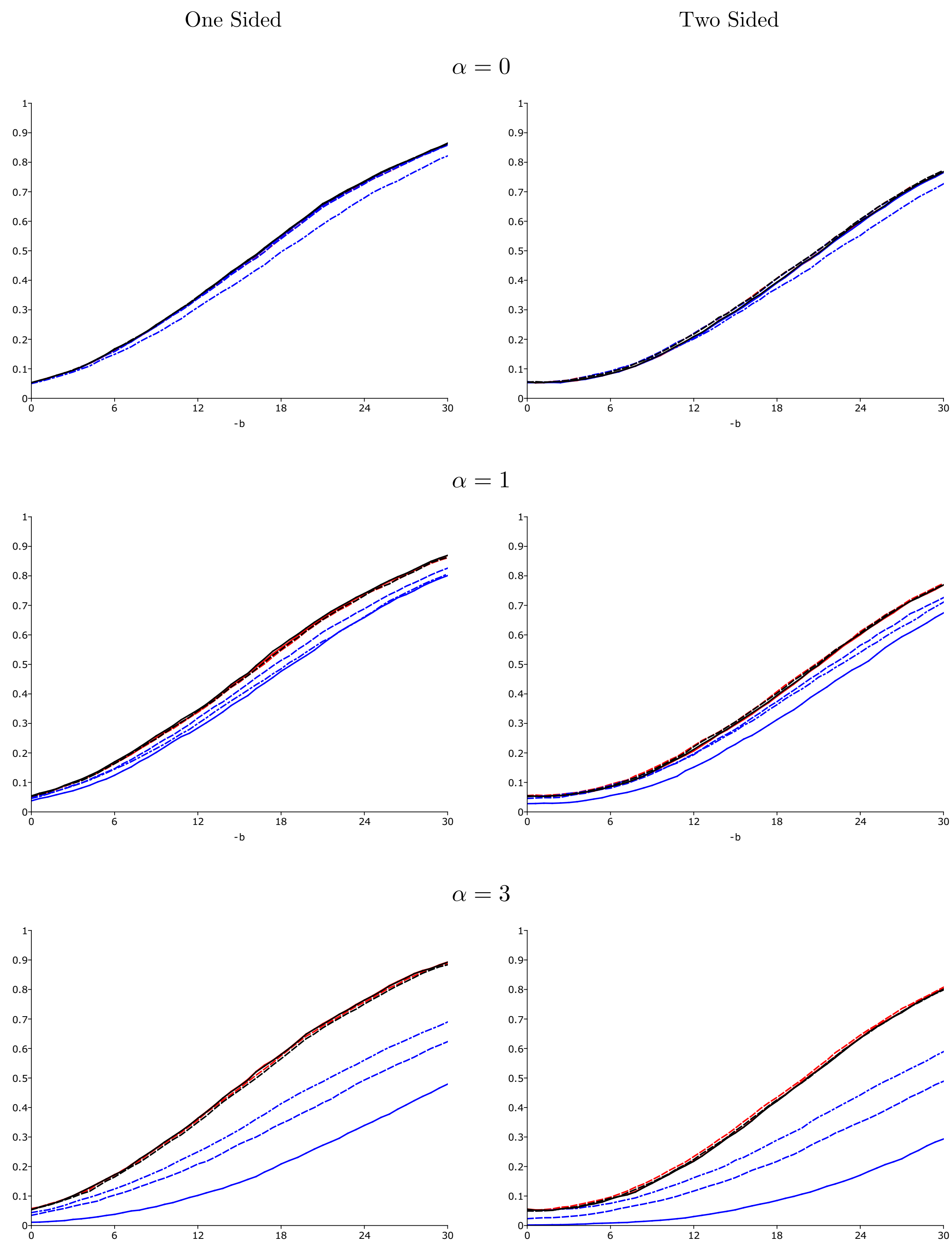
$\alpha = 3$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

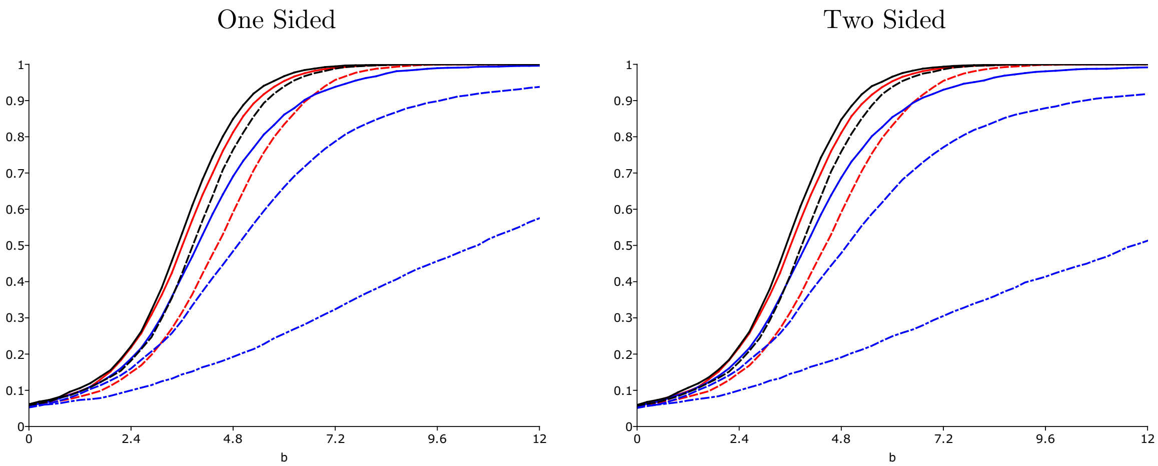
Figure S.9: Finite Sample Power, Left Tail, $n = 250$, $\delta = -0.95$, $c = 50$



$IVX_{0.95}$: ———, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ———, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ———, $U_{0.75}$: – –

Figure S.10: Finite Sample Power, Right Tail, $n = 1000$, $\delta = -0.95$, $c = 0$



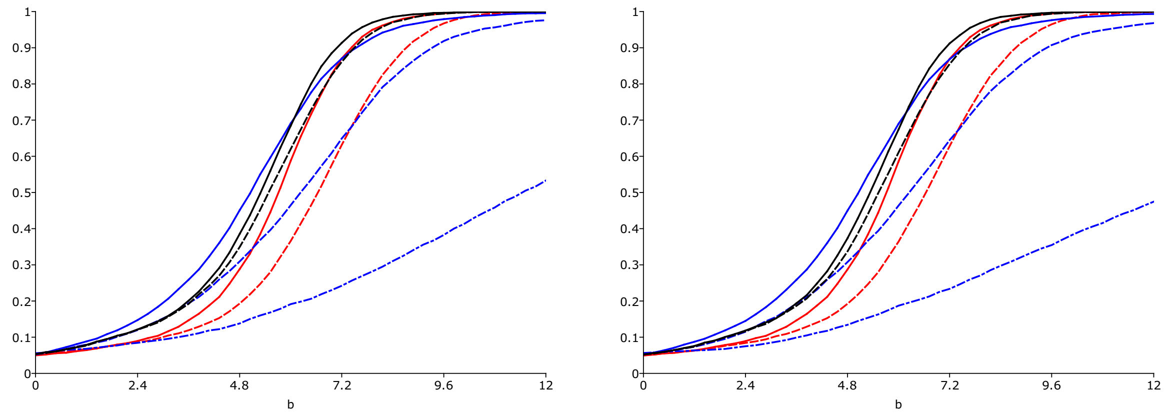
$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

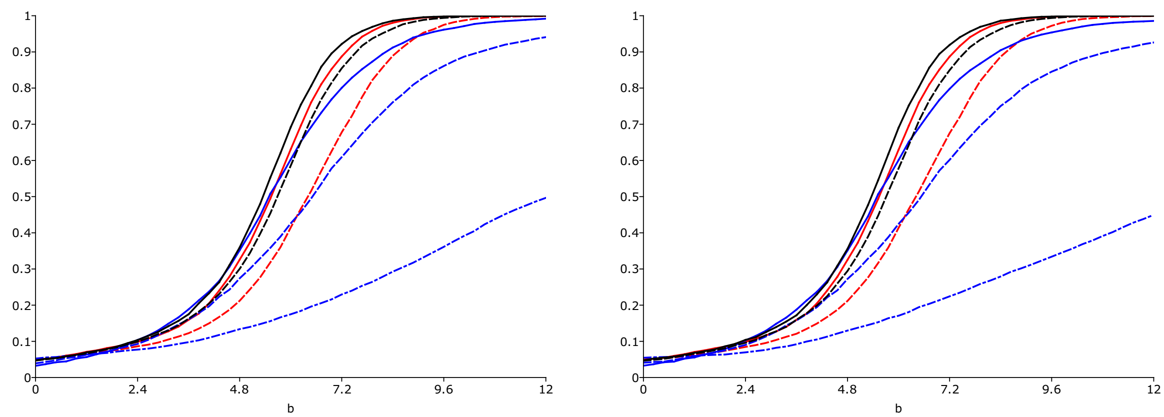
Figure S.11: Finite Sample Power, Right Tail, $n = 1000$, $\delta = -0.95$, $c = 2$

One Sided | Two Sided

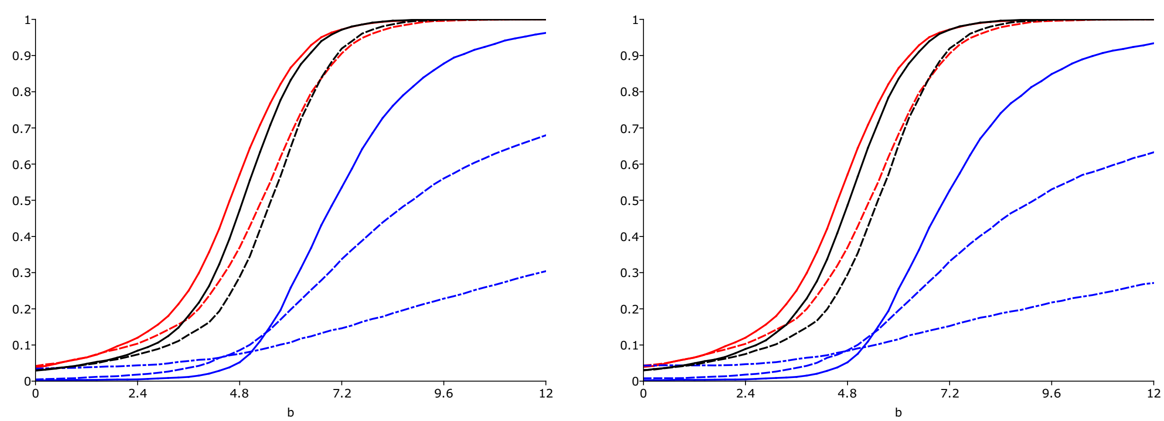
$\alpha = 0$



$\alpha = 1$



$\alpha = 3$



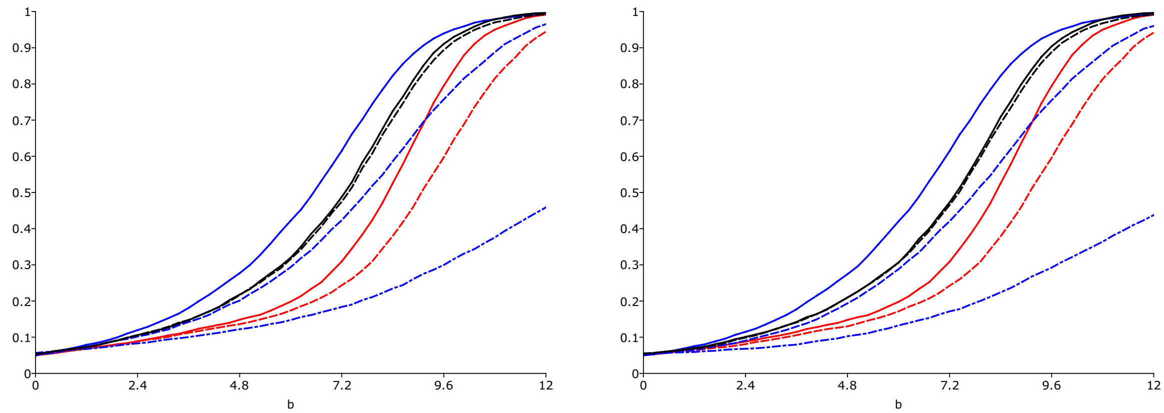
$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

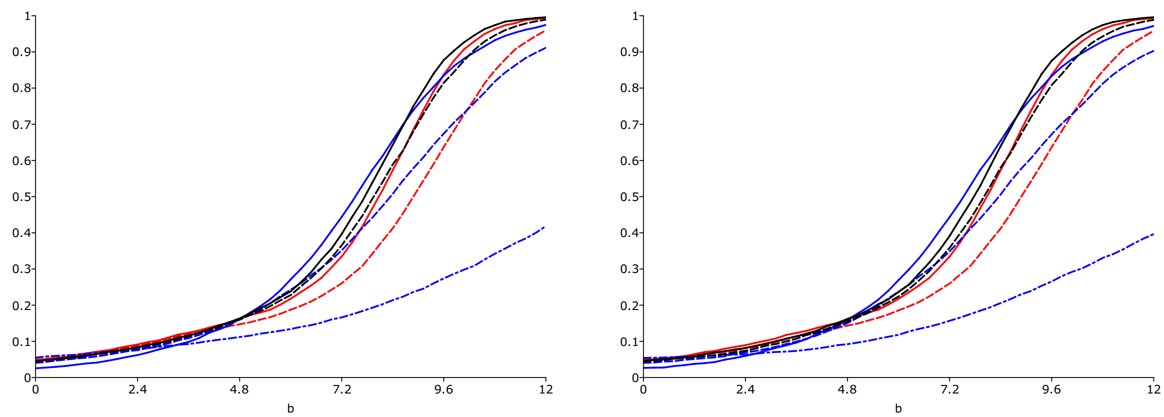
Figure S.12: Finite Sample Power, Right Tail, $n = 1000$, $\delta = -0.95$, $c = 5$

One Sided | Two Sided

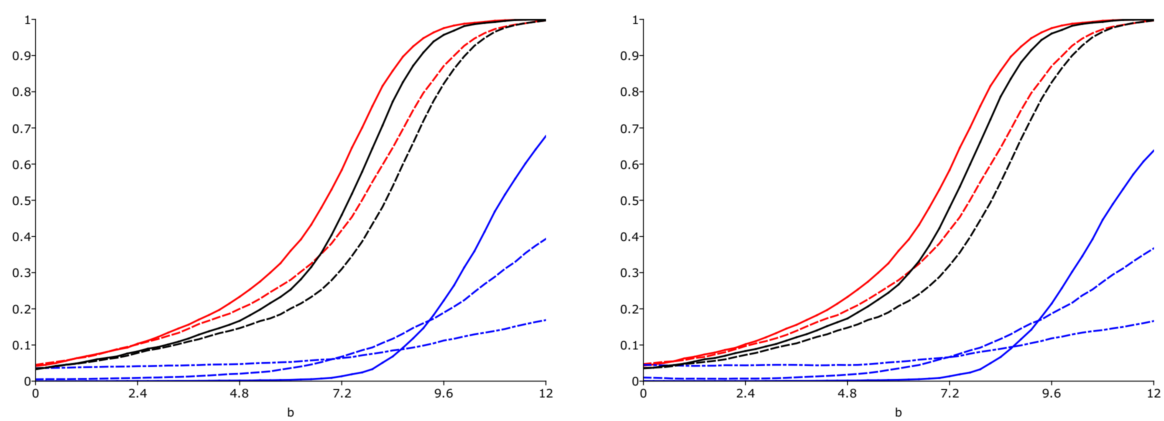
$\alpha = 0$



$\alpha = 1$



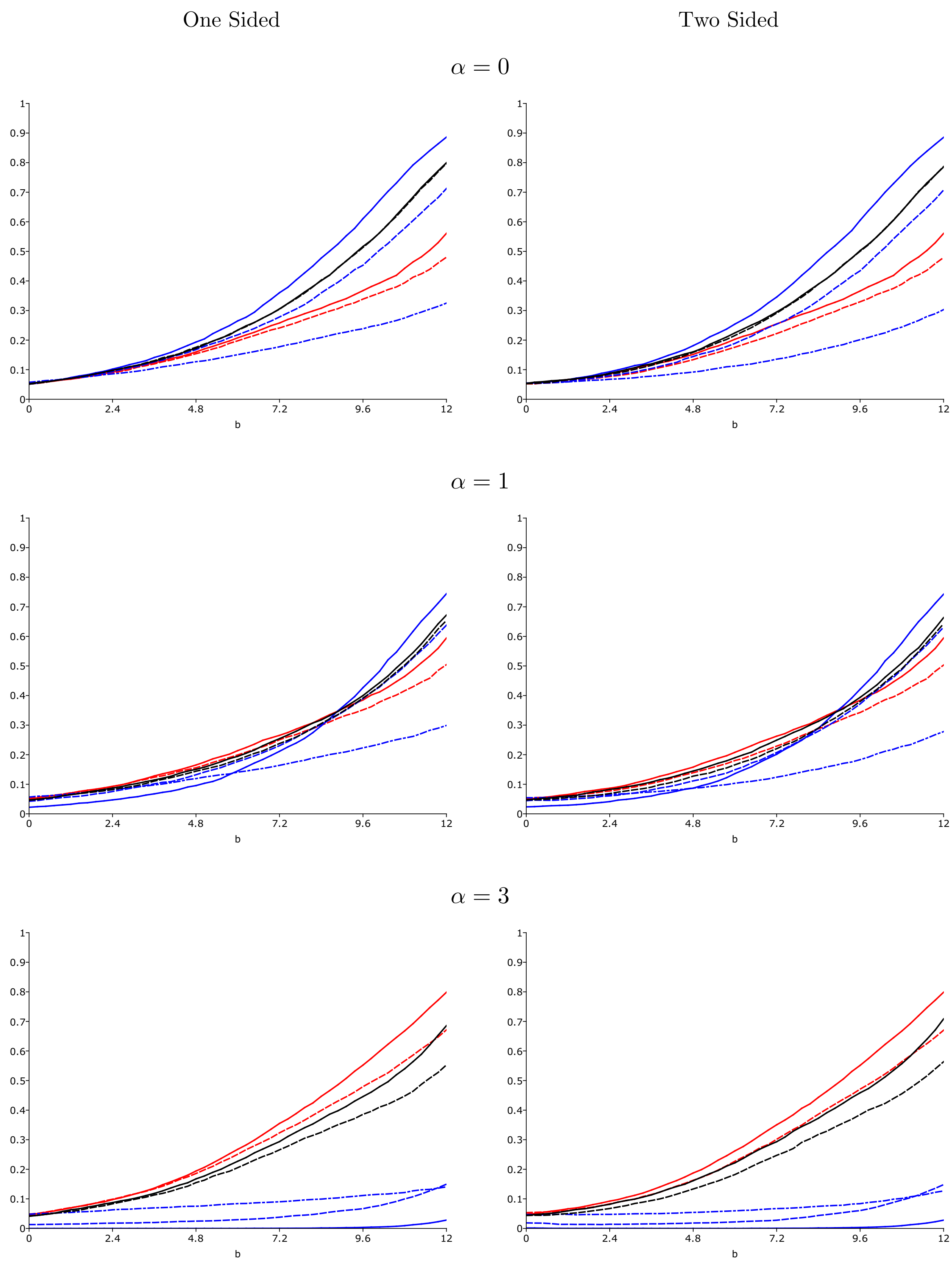
$\alpha = 3$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

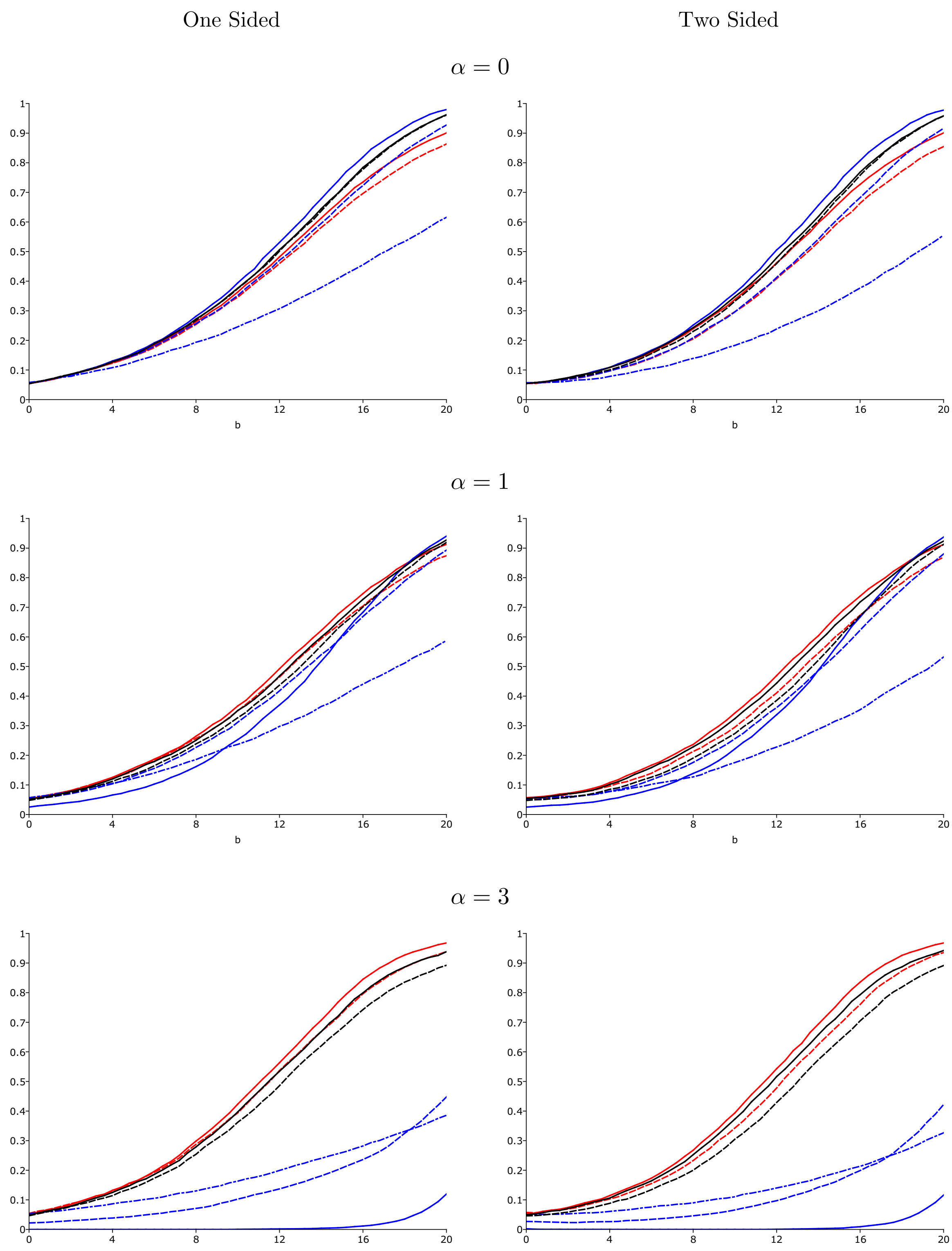
Figure S.13: Finite Sample Power, Right Tail, $n = 1000$, $\delta = -0.95$, $c = 10$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

Figure S.14: Finite Sample Power, Right Tail, $n = 1000$, $\delta = -0.95$, $c = 20$



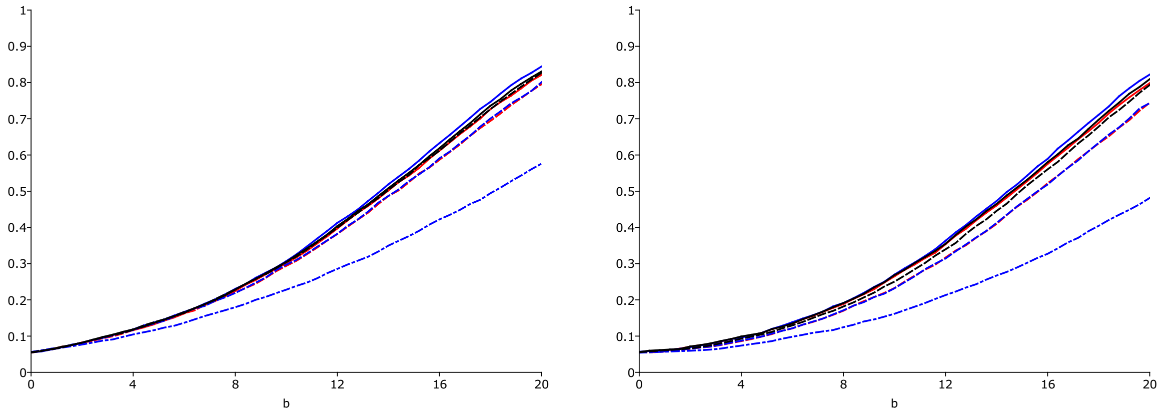
$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

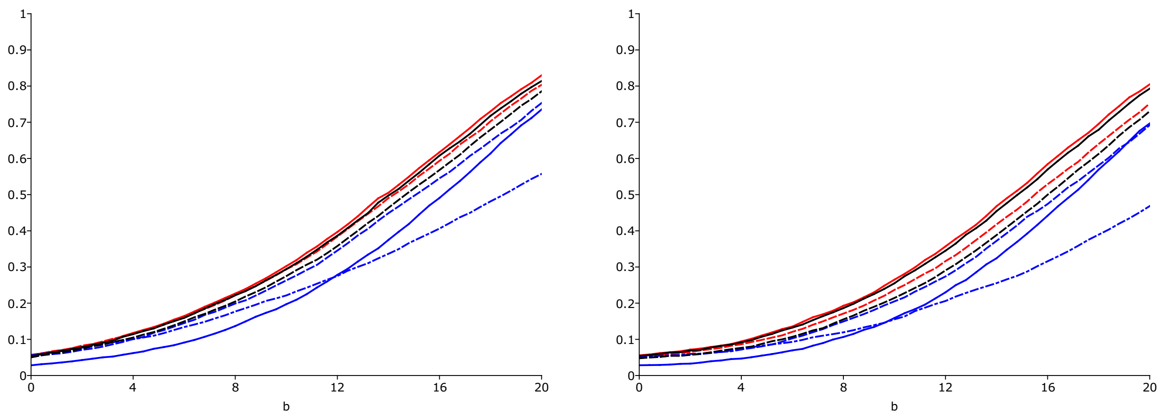
Figure S.15: Finite Sample Power, Right Tail, $n = 1000$, $\delta = -0.95$, $c = 30$

One Sided | Two Sided

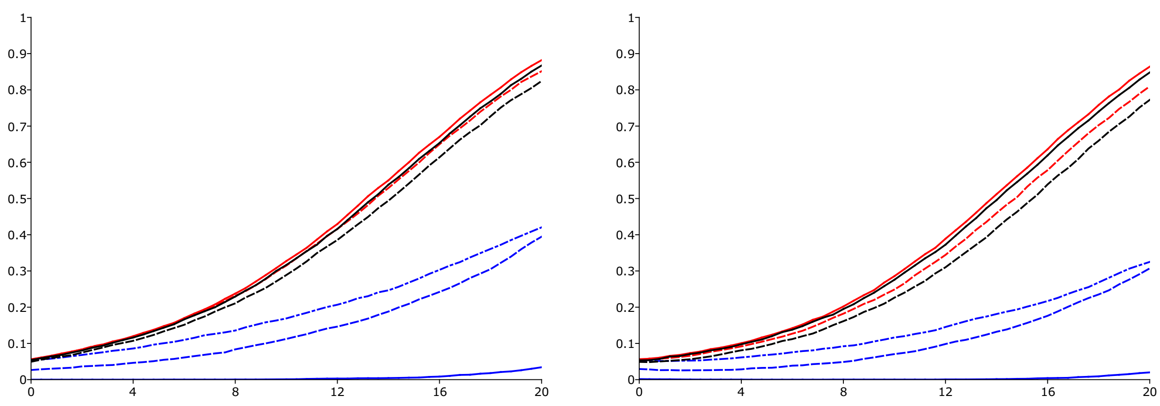
$\alpha = 0$



$\alpha = 1$



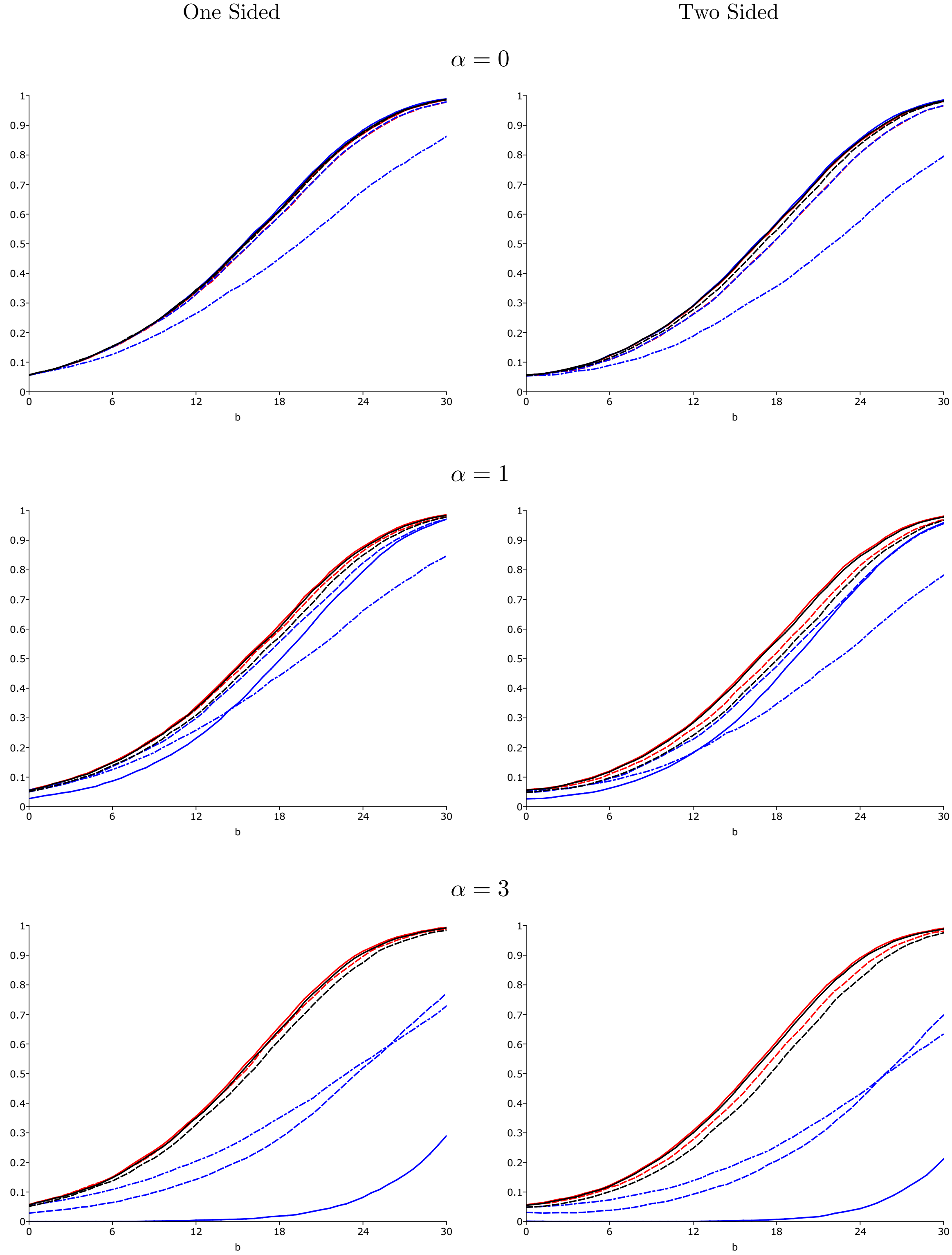
$\alpha = 3$



$IVX_{0.95}$: , $IVX_{0.75}$: , $IVX_{0.50}$: , $IVX^{\dagger}_{0.95}$: , $IVX^{\dagger}_{0.75}$: ,

$U_{0.95}$: , $U_{0.75}$:

Figure S.16: Finite Sample Power, Right Tail, $n = 1000$, $\delta = -0.95$, $c = 40$



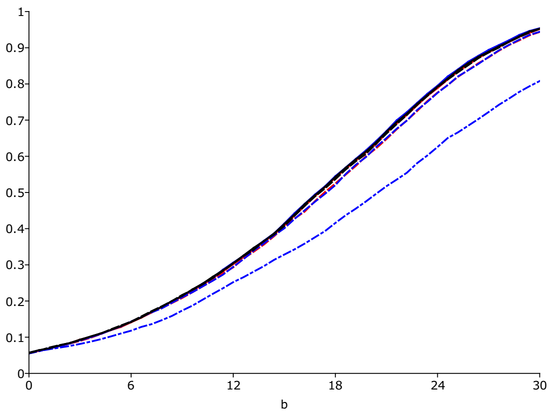
$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

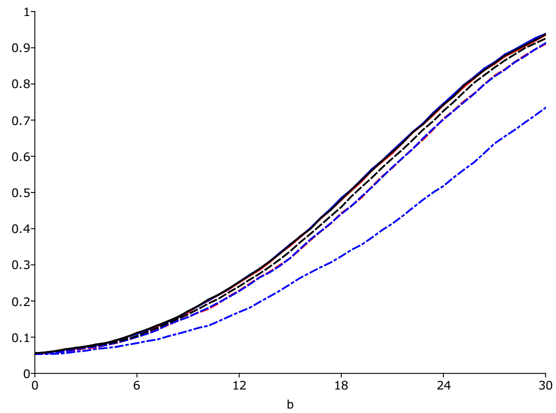
$U_{0.95}$: ——, $U_{0.75}$: – –

Figure S.17: Finite Sample Power, Right Tail, $n = 1000$, $\delta = -0.95$, $c = 50$

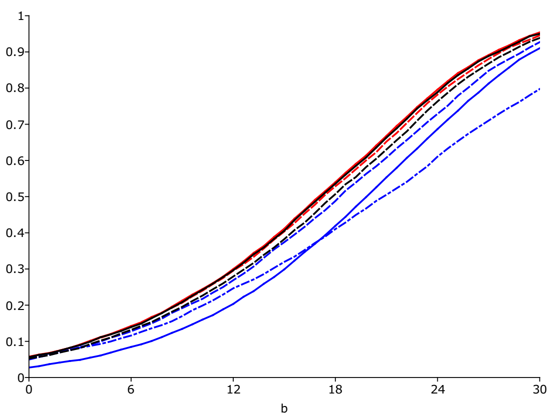
One Sided | Two Sided

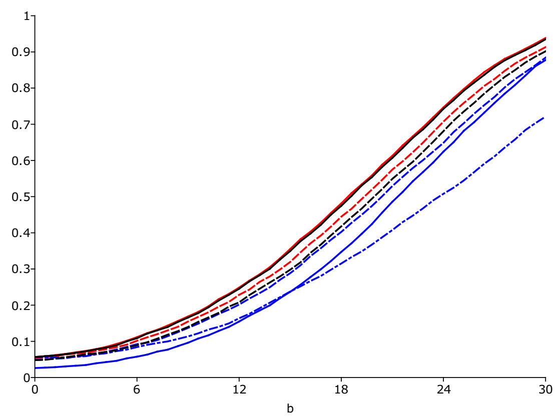
$\alpha = 0$



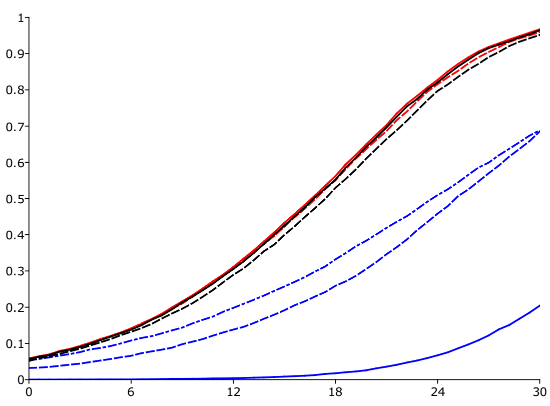


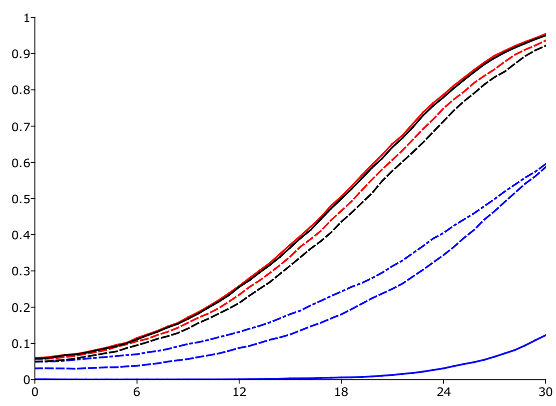
$\alpha = 1$





$\alpha = 3$

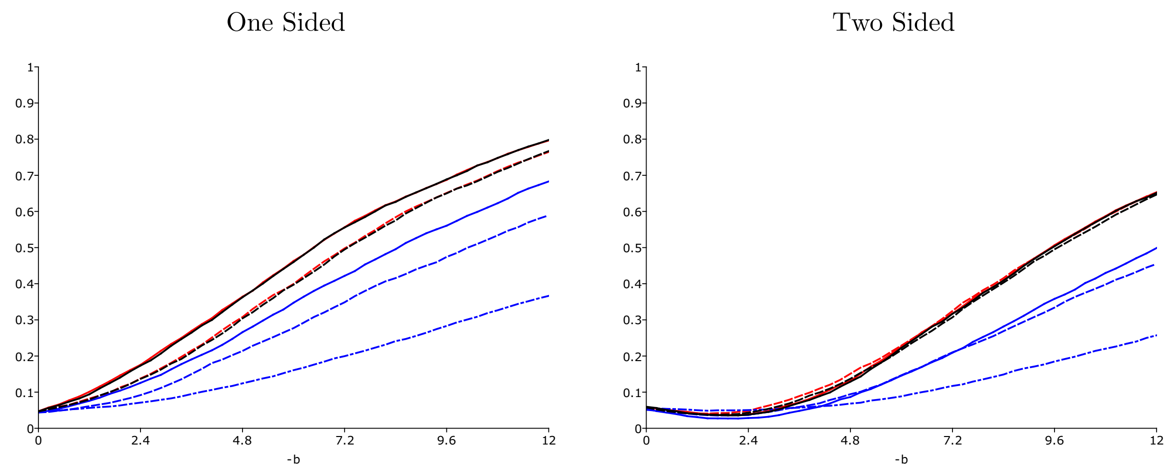




$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

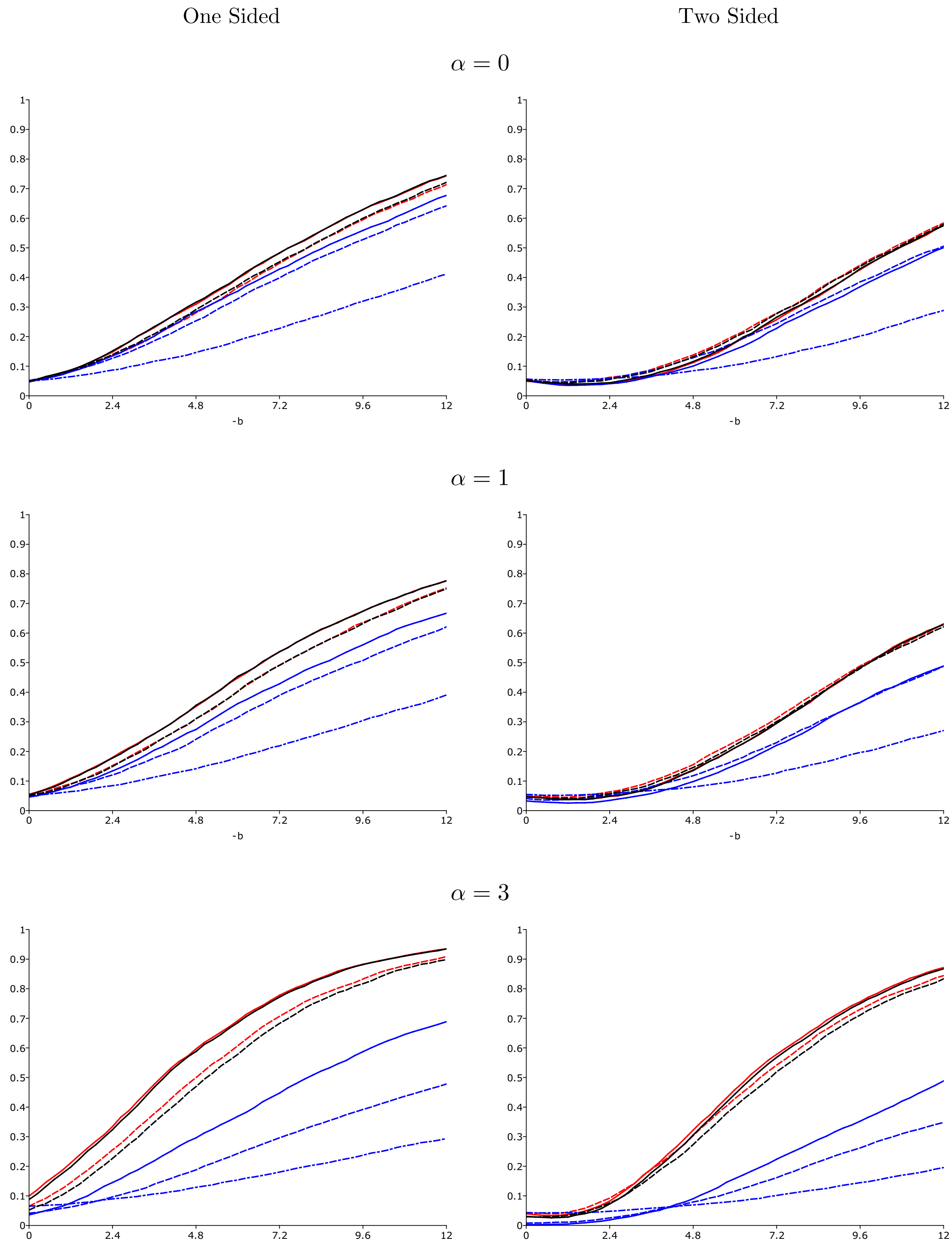
Figure S.18: Finite Sample Power, Left Tail, $n = 1000$, $\delta = -0.95$, $c = 0$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

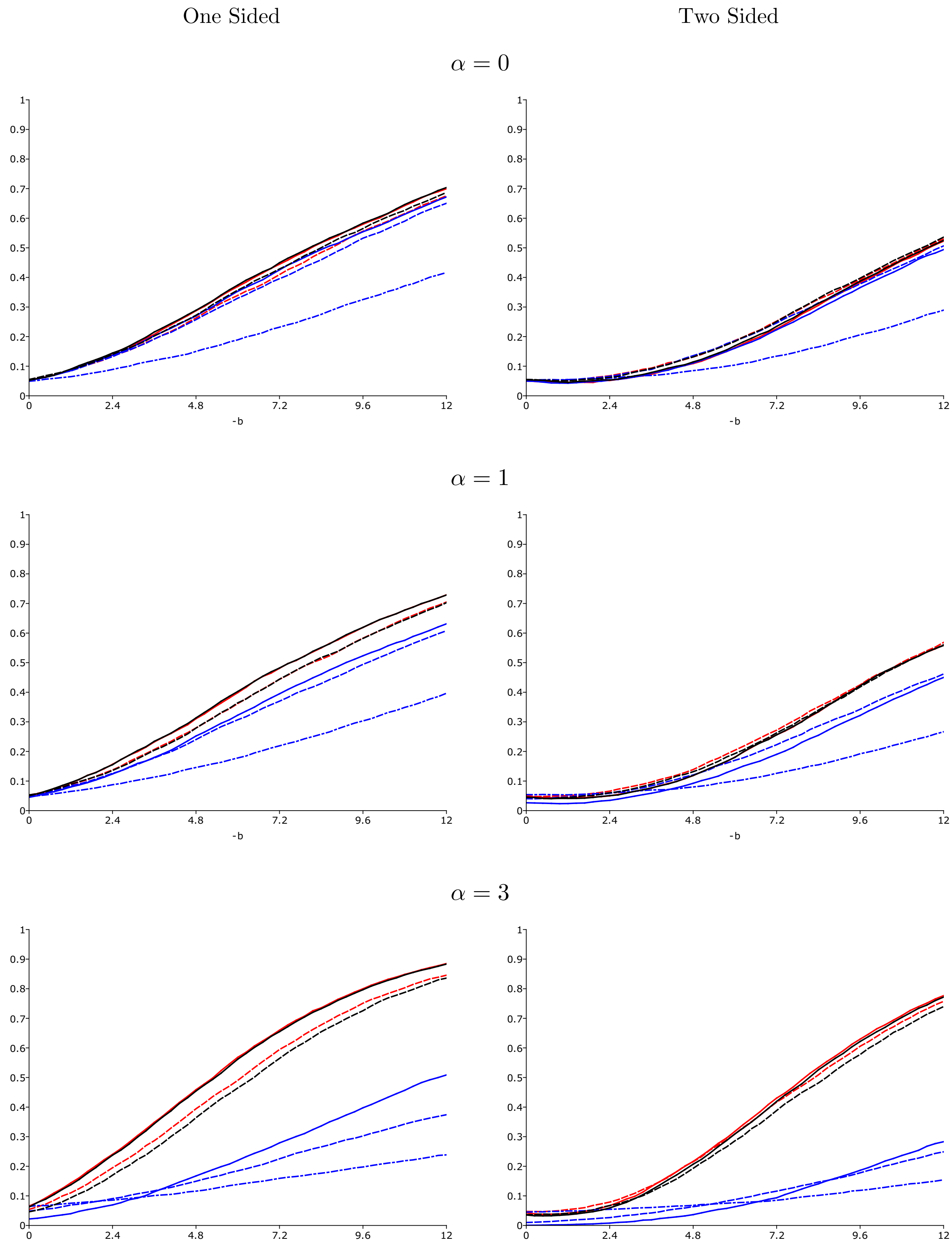
Figure S.19: Finite Sample Power, Left Tail, $n = 1000$, $\delta = -0.95$, $c = 2$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

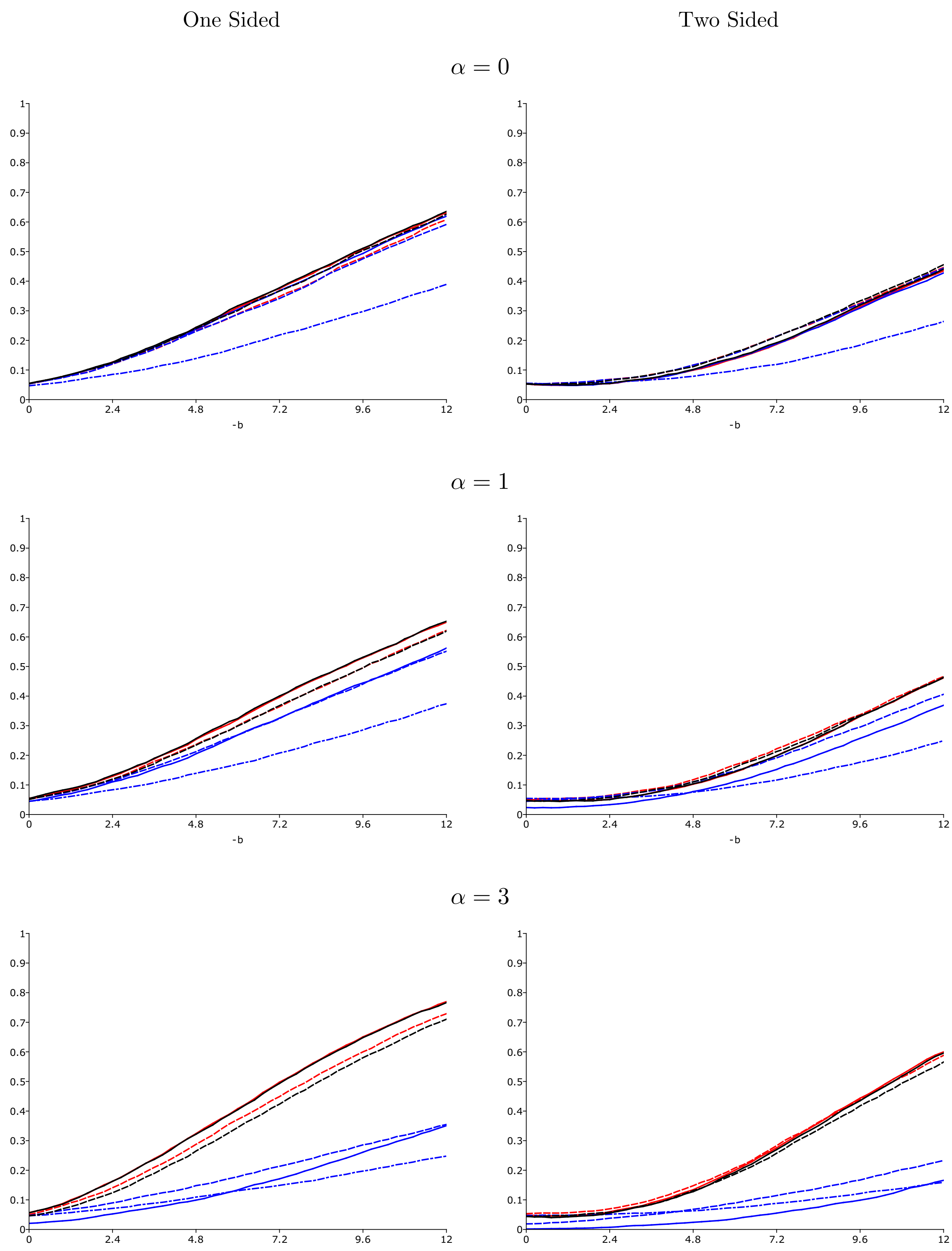
Figure S.20: Finite Sample Power, Left Tail, $n = 1000$, $\delta = -0.95$, $c = 5$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

Figure S.21: Finite Sample Power, Left Tail, $n = 1000$, $\delta = -0.95$, $c = 10$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

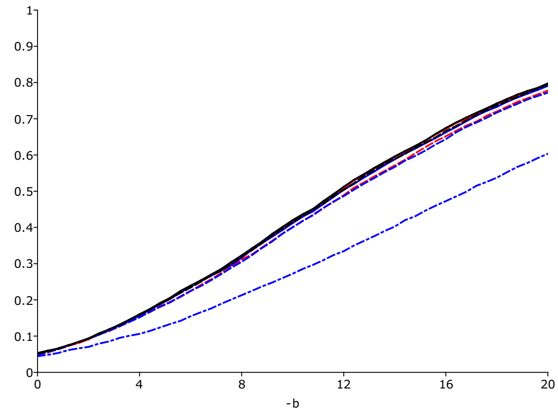
$U_{0.95}$: ——, $U_{0.75}$: – –

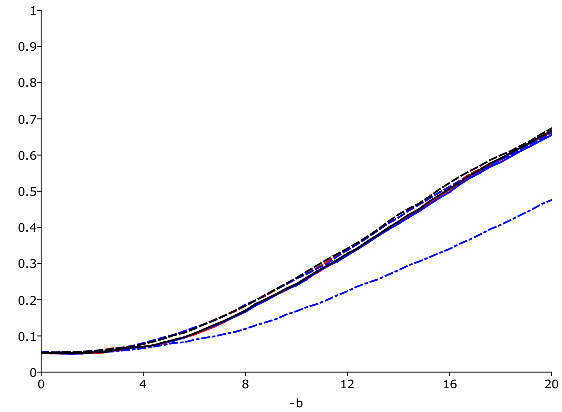
Figure S.22: Finite Sample Power, Left Tail, $n = 1000$, $\delta = -0.95$, $c = 20$

One Sided

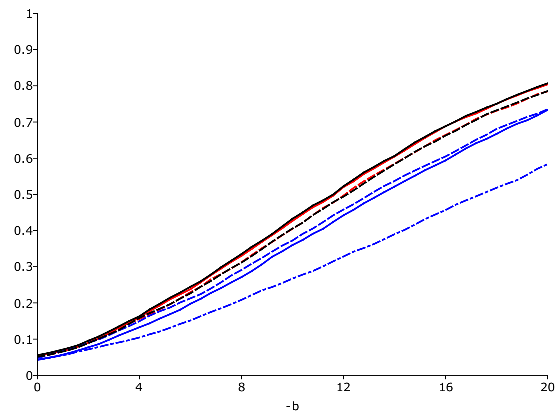
Two Sided

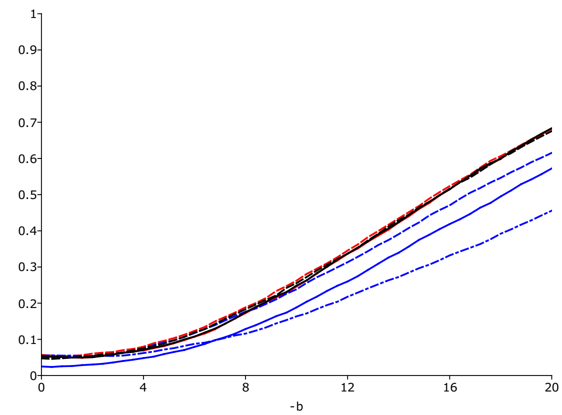
$\alpha = 0$



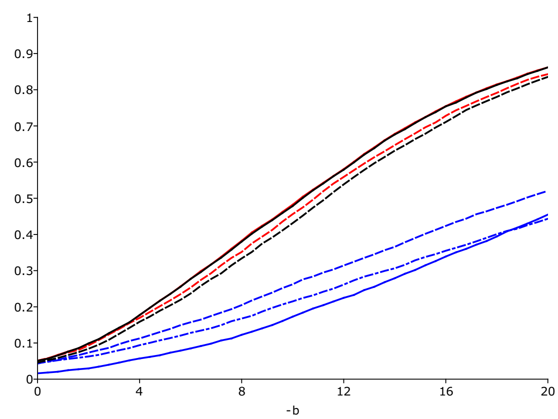


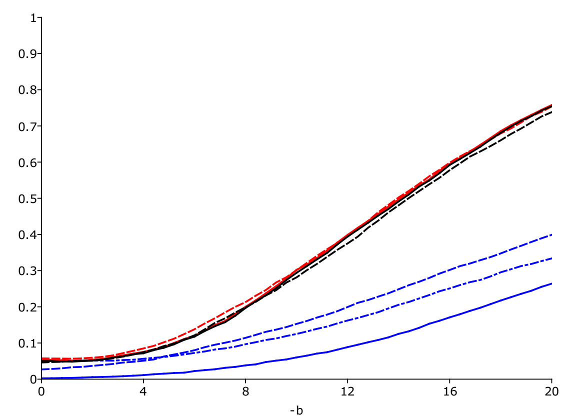
$\alpha = 1$





$\alpha = 3$

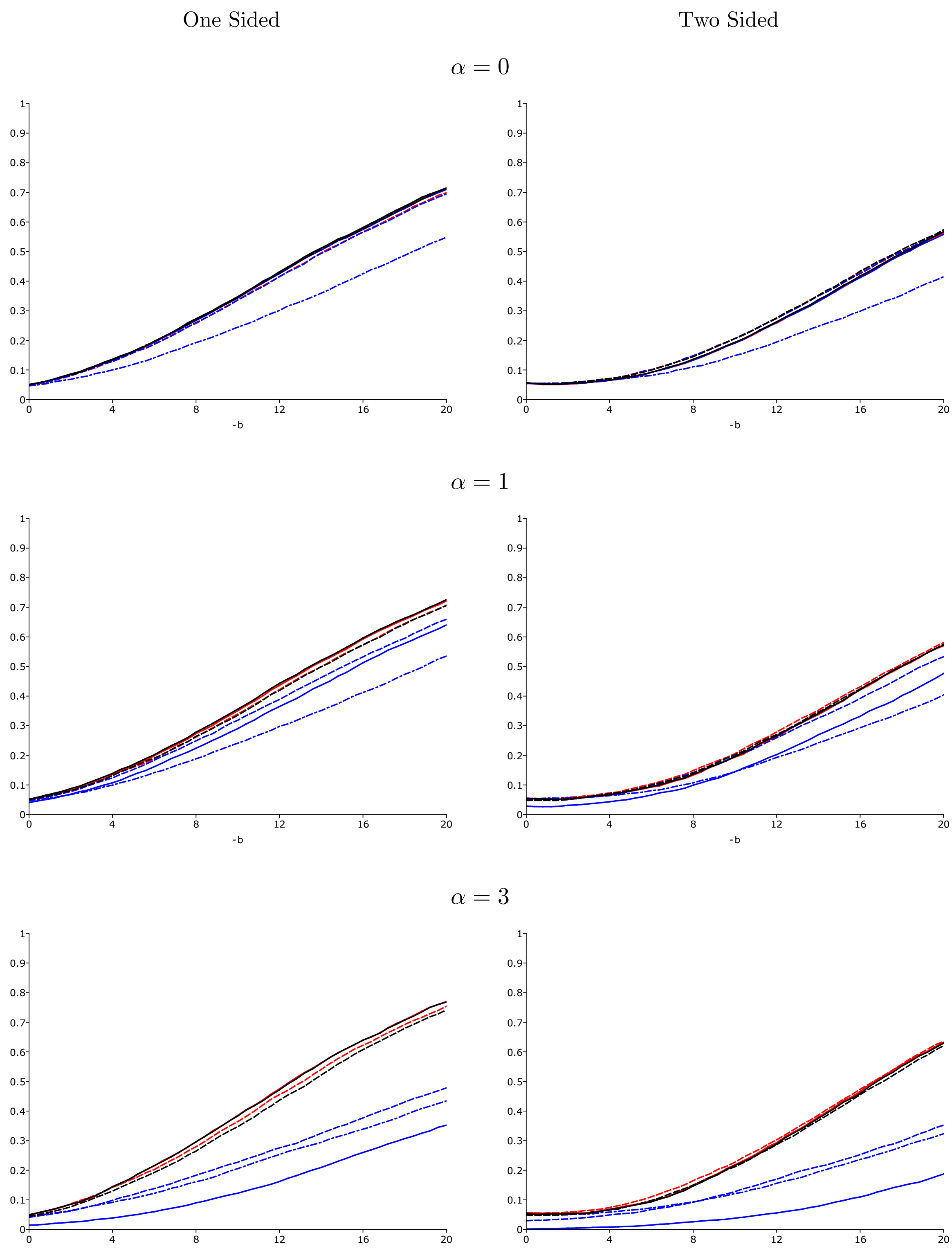




$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

Figure S.23: Finite Sample Power, Left Tail, $n = 1000$, $\delta = -0.95$, $c = 30$



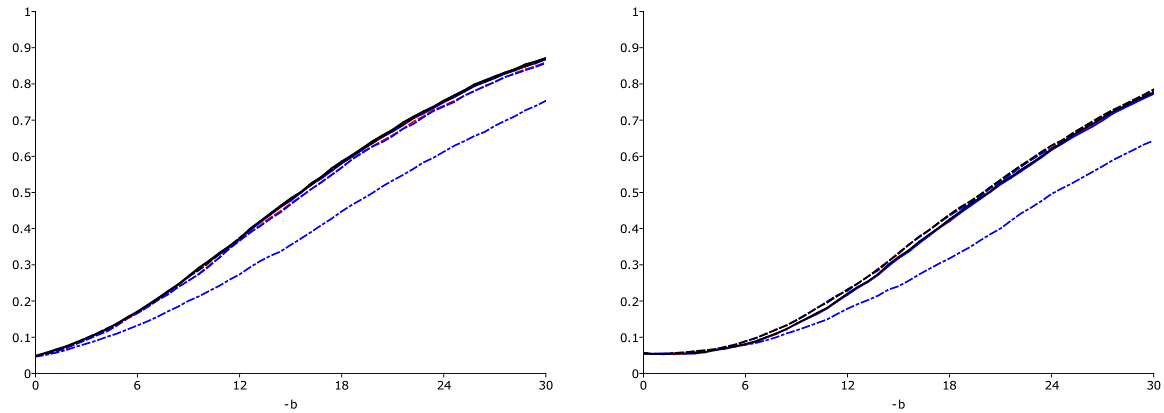
$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

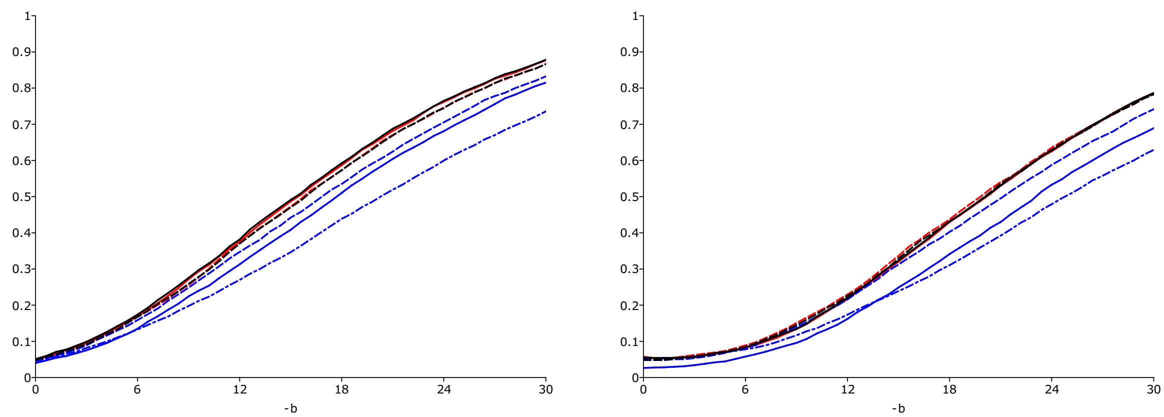
Figure S.24: Finite Sample Power, Left Tail, $n = 1000$, $\delta = -0.95$, $c = 40$

One Sided | Two Sided

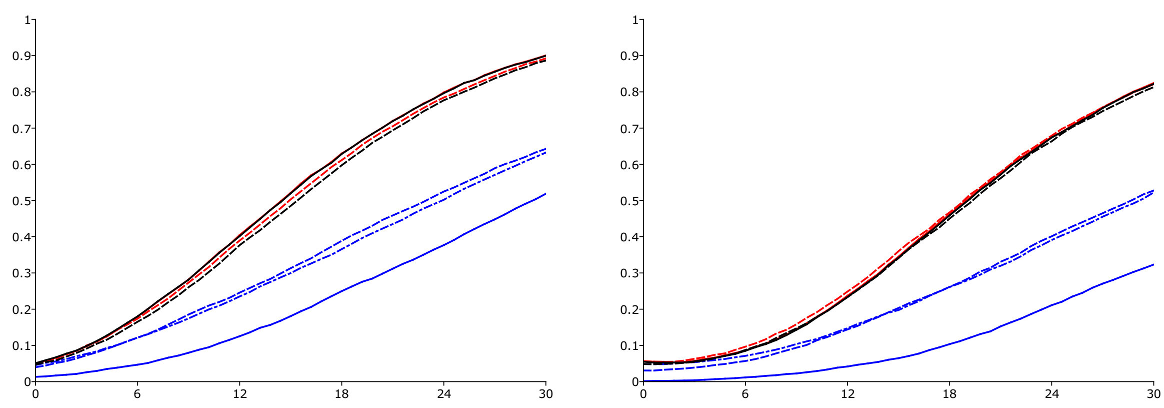
$\alpha = 0$



$\alpha = 1$



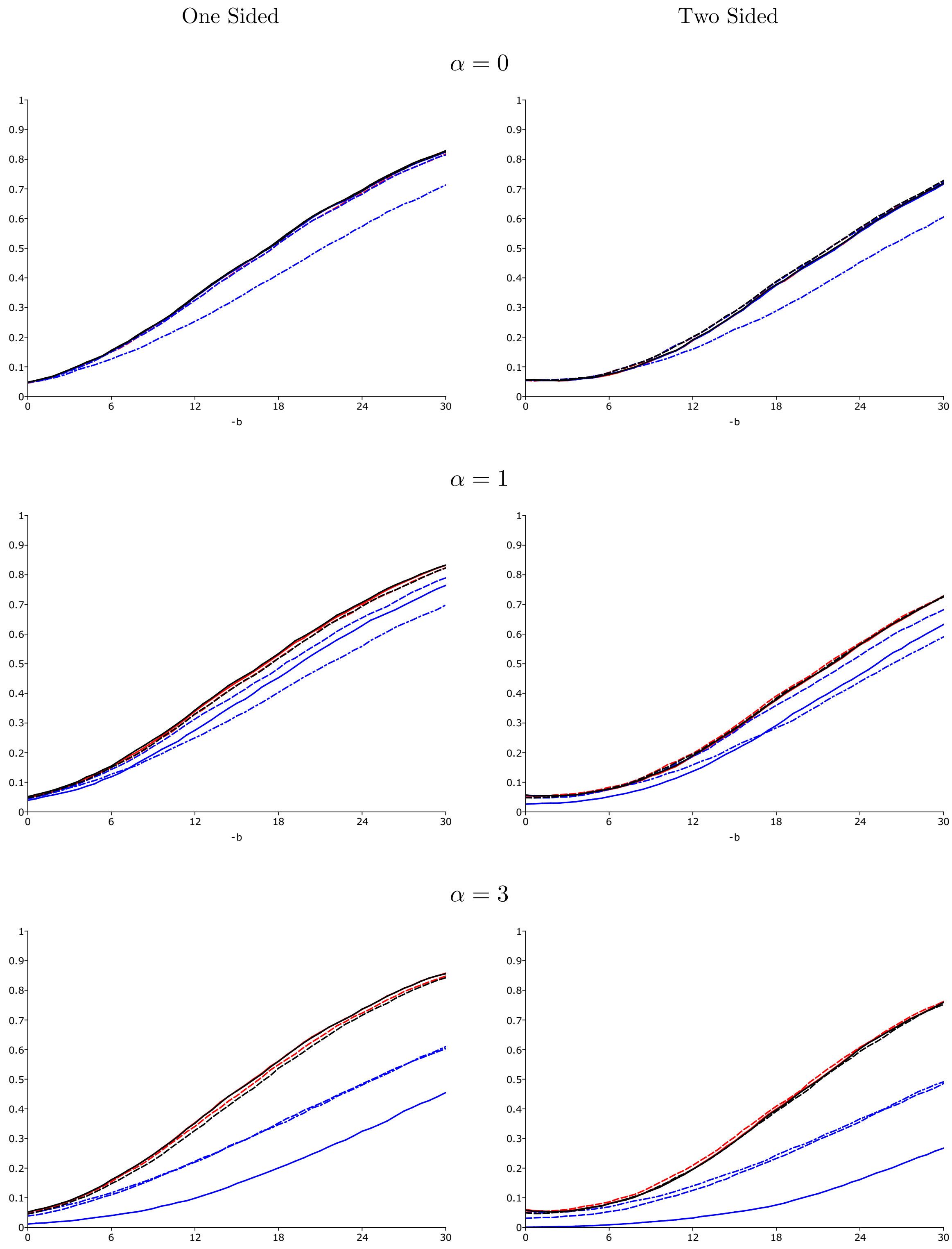
$\alpha = 3$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –

Figure S.25: Finite Sample Power, Left Tail, $n = 1000$, $\delta = -0.95$, $c = 50$



$IVX_{0.95}$: ——, $IVX_{0.75}$: – –, $IVX_{0.50}$: – · –, $IVX^{\dagger}_{0.95}$: ——, $IVX^{\dagger}_{0.75}$: – –,

$U_{0.95}$: ——, $U_{0.75}$: – –