Robust Econometric Inference for Stock Return Predictability

Alexandros Kostakis
University of Manchester

Tassos Magdalinos
University of Southampton

Michalis P. Stamatogiannis
University of Bath

We are grateful to Geert Bekaert (Editor), David Hirshleifer (Executive Editor) and four anonymous referees for their extensive comments and helpful suggestions. We would also like to thank Karim Abadir, Paul Bekker, Walter Distaso, Christoph Hanck, Erik Hjalmarsson, Jan Jacobs, Ioannis Kasparis, Lynda Khalaf, Timos Papadopoulos, Peter Phillips, Jean-Yves Pitarakis, Marno Verbeek (discussant), Michael Wolf (discussant) as well as conference participants at Netherlands Econometric Study group 2011 (Rotterdam), ESEM 2011 (Oslo), Statistical Week 2011 (Leipzig), EFA 2012 (Copenhagen), RES 2014 (Manchester), Marie Curie ITN on Risk Management and Risk Reporting (Konstanz) and seminar participants at Aston Business School, Bath, Loughborough, Manchester Business School, Queen Mary (London) and Southampton for their insightful comments and suggestions. Magdalinos gratefully acknowledges financial support by the ESRC grant RES-061-25-0517.
Send correspondence to Tassos Magdalinos, Department of Economics, University of Southampton, SO17 1BJ, Southampton, UK. Phone: +44 (0) 2380593175. E-mail: a.magdalinos@soton.ac.uk.

Robust Econometric Inference for Stock Return Predictability

Abstract

This study examines stock return predictability via lagged financial variables with unknown stochastic properties. We conduct a battery of predictability tests for US stock returns during the 1927–2012 period, proposing a novel testing procedure which: *i*) robustifies inference to the degree of persistence of the employed regressors, *ii*) accommodates testing the joint predictive ability of financial variables in multiple regression, *iii*) is easy to implement as it is based on a linear estimation procedure and *iv*) can be also used for long-horizon predictability tests. We provide some evidence in favor of short-horizon predictability in the full sample period. Nevertheless, this evidence almost entirely disappears in the post–1952 period. Moreover, predictability becomes weaker, not stronger, as the predictive horizon increases.

Keywords: Stock returns; Predictability; Persistent regressors; Robust inference.

A fundamental issue in finance is whether future stock returns are predictable using publicly available information (see Fama, 1970). The seminal studies of Keim and Stambaugh (1986), Fama and French (1988) and Campbell and Shiller (1988) empirically demonstrated that certain financial variables have significant predictive ability over future stock returns. Fama (1991) interpreted these findings as evidence of time-varying risk premia rather than evidence against market efficiency. Despite the significant volume of subsequent research, the predictability debate still remains unsettled (see Ang and Bekaert, 2007, for an insightful discussion). On the one hand, Lettau and Ludvigson (2001, p. 842) state that "it is now widely accepted that excess returns are predictable by variables such as dividend-price ratios, earning-price ratios, dividend-earnings ratios and an assortment of other financial indicators". But many remain sceptical, claiming that the "profession has yet to find some variable that has meaningful and robust empirical equity premium forecasting power both in-sample and out-of-sample" (Welch and Goyal, 2008, p. 1505).

Empirical support of arguments in favor of or against predictability crucially relies on inference from predictive regressions, hence the size and power of the employed hypothesis tests assume fundamental importance. A series of recent studies, reviewed in Campbell and Yogo (2006) (hereafter CY), recognize that the most common problem undermining confidence in the reliability of predictability tests is the uncertainty about the (unobservable) time series properties of the predictor variables and, in particular, their degree of persistence. Regardless of one's prior beliefs on their order of integration, it is well documented that most of the variables used in predictive regressions are highly persistent with autoregressive roots extremely close to unity (see CY, and Welch and Goyal, 2008). This empirical fact casts doubt on the validity of standard t-tests based on least squares regressions (see Cavanagh, Elliott and Stock, 1995, and Torous, Valkanov and Yan, 2004). As Stambaugh (1999) has convincingly shown, this problem is exacerbated if, additionally, the innovations of the predictor are highly correlated with the innovations of the returns, i.e., when the predictive regressor is endogenous. Endogeneity is a typical feature of commonly used predictors, such as price-scaled ratios. Since regression estimators and tests have fundamentally different properties in the presence of persistent and endogenous predictors, confidence in the reliability of predictability tests is undermined, as the quality of inference is conditional upon correct specification of the predictors' time series properties.

Acknowledging the uncertainty regarding the degree of predictive variables' persistence, a strand of the literature suggests modelling these variables as local-to-unity processes (see inter alia Lanne, 2002, Valkanov, 2003, Torous et al., 2004, CY, Jansson and Moreira, 2006, and Hjalmarsson, 2011). These processes assume the form of a first-order autoregression with root $\rho = 1 + c/n$, approaching a random walk as the sample size n increases to infinity. While providing flexibility in modelling, the use of explanatory variables that exhibit persistence without necessarily being random walks in finite samples raises serious technical complications. Since standard cointegration methods cannot accommodate the presence of local-to-unity roots in predictive regressions, Cavanagh et al. (1995), Torous et al. (2004), CY, and Hjalmarsson (2011) have employed methods based on inverting the non-pivotal limit distribution of the t-statistic and constructing Bonferroni-type confidence intervals for the nuisance parameter c. This is the current state of the art methodology for testing the predictability of stock returns with highly persistent regressors.

Practical implementation of the above methodology presents two main drawbacks. First, the method is invalid if the regressor contains stationary or near stationary components; the validity of the method requires each predictor to be at least as persistent as a local-to-unity process, a restrictive assumption that cannot be empirically tested. Second, due to the problems associated with the construction of multidimensional confidence intervals for c, the methodology is restricted to the case of a scalar regressor, i.e., a single predictive variable. This imposes a severe restriction, since the joint predictability by combinations of financial variables cannot be tested. The above framework can only accommodate testing the predictive power of each financial variable in isolation, which may result in loss of information through omitted variables. These limitations have also been indicated by Ang and Bekaert (2007, footnote 3). We build upon this strand of the literature by proposing a methodology that successfully overcomes these limitations.

In recent work, Phillips and Magdalinos (2009) provide a framework of limit theory that can be used to validate inference in models with regressors exhibiting very general time series characteristics. Endogeneity is successfully removed by means of a data filtering procedure called IVX estimation. The key idea behind the method is the explicit control of the degree of persistence of data-filtered IVX instruments, restricted within the class of near stationary processes. In this study, we prove that in the context of multivariate predictive regressions, the IVX approach yields standard chi-squared asymptotic inference for testing general restrictions on predictive variables with degree of persistence covering the entire range from stationarity of stable autoregressions to pure nonstationarity of unit root processes. The robustness of the IVX approach should alleviate practical concerns about the quality of inference under possible misspecification of the time series properties of the predictive regressors. The dimensionality of the system of predictive regressions is of considerable practical importance too, since the IVX methodology enables the assessment of the joint predictive power of various combinations of regressors.[1] In summary, our study introduces and implements a testing procedure that resolves two important outstanding issues in the predictability literature: i) robustness with respect to the time series properties of the predictors and ii) joint testing in systems of predictive equations. Furthermore, we show that this testing procedure is also applicable to long-horizon predictive regressions and we develop the relevant statistic.

We implement the proposed methodology by conducting a battery of short- and long-horizon predictability tests for US stock returns during the 1927–2012 period, using a set of commonly employed variables. We focus on in-sample predictability tests since the proposed methodology aims to robustify in-sample inference with respect to regressors' unknown time series properties. In univariate tests, we find significant predictive ability with respect to 1-period ahead excess market returns for the earnings-price and book-to-market value ratios as well as net equity expansion. However, this evidence almost entirely disappears in the post–1952 period. Only the consumption-wealth ratio is found to be strongly significant in this subperiod. Our multivariate tests show that the combination of the earnings-price ratio and T-bill rate is highly significant and robust to the choice of data frequency and examined period. Finally, with respect to long-horizon tests, we find that, if anything, predictability generally becomes weaker, not stronger, as the horizon increases. Only the consumption-wealth ratio remains strongly significant for all horizons examined.

The rest of this study is organized as follows. Section 1 presents the IVX methodology

[1] It should be noted that the iterative procedure of Amihud, Hurvich and Wang (2009) also accommodates multiple predictors under the restriction that these are stationary. Moreover, the recent contribution of Kelly and Pruitt (2013) also utilizes a multivariate system of predictive regressions. However, their focus is on extracting information regarding aggregate expected returns and dividend growth from the cross-section of price-dividend ratios using the present value relationship that has been employed for predictability tests *inter alia* by Lettau and Ludvigson (2005), Cochrane (2008), Lettau and van Nieuwerburgh (2008) and van Binsbergen and Koijen (2010).

in systems of predictive regressions, while a comprehensive Monte Carlo study for the finite-sample properties of the derived test statistic is provided in Section 2. Section 3 presents the dataset used in the empirical analysis. In Section 4 we carry out short-horizon predictability tests for US stock returns, while in Section 5 we generalize the testing procedure to long-horizon predictive regressions, we examine the finite-sample properties of the corresponding statistic and we present the results from long-horizon predictability tests. Section 6 contains some concluding remarks. The Appendix outlines the large-sample distributional properties of IVX estimation and inference. The proofs are collected in the Online Appendix, which also contains a series of further results.

1. Robust inference for predictive regressions

This section develops an econometric methodology for testing stock return predictability that is robust to uncertainty over the stochastic properties of the financial variables used as potential predictors. Accommodating this uncertainty requires a modelling framework that encompasses all empirically relevant classes of autoregressive data generating mechanisms. To this end, we consider the following multivariate system of predictive regressions with regressors containing explanatory variables with arbitrary degree of persistence:

$$y_t = \mu + A x_{t-1} + \varepsilon_t, \tag{1}$$

$$x_t = R_n x_{t-1} + u_t, \tag{2}$$

where A is an $m \times r$ coefficient matrix and

$$R_n = I_r + \frac{C}{n^\alpha} \quad \text{for some } \alpha \geq 0, \tag{3}$$

and some matrix $C = \text{diag}(c_1, ..., c_r)$, where n denotes the sample size. The vector of predictive variables x_t in (2) exhibits a degree of persistence induced by the autoregressive matrix in (3) that belongs to one of the following persistence classes:

P(i) *Integrated regressors, if* $C = 0$ *or* $\alpha > 1$ *in (3).*

P(ii) *Local-to-unity regressors, if* $C \neq 0$ *and* $\alpha = 1$ *in (3).*

P(iii) *Near stationary regressors, if* $c_i < 0$ *for all* i *and* $\alpha \in (0,1)$ *in (3).*

P(iv) *Stationary regressors, if* $c_i < 0$ *for all* i *and* $\alpha = 0$ *in (3).*

The classes P(i)-P(iv) include predictors with very general time series characteristics varying from purely stationary to purely non-stationary processes and accommodating all intermediate persistence regimes. The predictive regression system may be initialized at some x_0 that could be any fixed constant vector or a random process satisfying $\|x_0(n)\| = o_p\left(n^{1/2}\right)$ when $\alpha \geq 1$ or $\alpha = 0$ and $\|x_0(n)\| = o_p\left(n^{\alpha/2}\right)$ when $\alpha \in (0,1)$.

Estimators and test statistics for conducting inference on the matrix A have very different properties according to the classification of the predictor process in (2) into one of the above persistence classes. Standard tests are asymptotically valid only within each class P(i)–P(iv) and misspecification of the degree of predictor persistence may lead to severe size distortions, particularly in the presence of endogeneity, i.e., correlation between the innovations ε_t and u_t of the predictive regression system (1)–(2) (see Elliott, 1998).[2] CY have partly addressed the problem for univariate predictive regressions ($m = r = 1$ in (1)–(2)) by inverting the limit distribution of the t-statistic under a local-to-unity regime P(ii) and using the Bonferroni inequality to construct confidence intervals which are asymptotically valid under P(i) or P(ii). However, the CY method loses its asymptotic validity for predictors that lie closer to the stationary region than local-to-unity time series. Such predictors can be modelled either as local-to-unity processes with c_i in (3) being large in absolute value (Phillips, 1987) or, more formally, as belonging to the class P(iii) of near stationary processes established by Phillips and Magdalinos (2007) and extended to multivariate systems of regression equations by Magdalinos and Phillips (2009).

We provide valid inference on A when there is no a priori knowledge of whether x_t belongs to class P(i), P(ii), P(iii) or P(iv). Our methodology for achieving robust inference is based on the IVX instrumentation procedure proposed by Phillips and Magdalinos (2009). The intuition

[2] In general, long-run endogeneity cannot be removed by standard cointegration methods such as the fully modified least squares estimation of Phillips and Hansen (1990) or the approaches of Saikkonen (1991) and Stock and Watson (1993) that apply when the regressor is a pure random walk ($c = 0$). As pointed out by Elliott (1998), such endogeneity corrected estimators lose their asymptotic mixed normality property under a local-to-unity regime and the associated hypothesis tests have a non-standard limit distribution, with the non-centrality parameter depending on the coefficient c of the local-to-unity root. Since c cannot be consistently estimated, the endogeneity cannot be removed, leading to asymptotically invalid predictability tests. Analogous problems arise when predictors exhibit a lower degree of persistence relative to local-to-unity processes, as is the case with the class of "near stationary" processes introduced by Phillips and Magdalinos (2007) as well as stationary autoregressive processes.

behind this procedure is to construct an instrumental variable whose degree of persistence we explicitly control. In this way, the inference problems arising due to the uncertainty regarding the persistence of the original regressor are avoided. Using the constructed instrument, one then performs a standard instrumental variable estimation. It turns out that the derived estimator follows asymptotically a mixed normal distribution, and hence the corresponding Wald statistic follows asymptotically a chi-squared distribution under the null, considerably simplifying inference.

To fix ideas, we construct near stationary instruments belonging to the class P(iii) by differencing the regressor x_t and constructing a new process according to an artificial autoregressive matrix with specified persistence degree. Despite the fact that the difference:

$$\Delta x_t = u_t + \frac{C}{n^\alpha} x_{t-1}$$

is not an innovation unless the regressor belongs to the class P(i) of integrated processes, it behaves asymptotically as an innovation after linear filtering by a matrix consisting of near stationary roots of the type P(iii). Choosing an artificial matrix:

$$R_{nz} = I_r + \frac{C_z}{n^\beta}, \quad \beta \in (0,1)\,, \quad C_z < 0, \tag{4}$$

IVX instruments $\tilde{z}_t$ are constructed as a first-order autoregressive process with autoregressive matrix R_{nz} and innovations Δx_t:

$$\tilde{z}_t = R_{nz}\tilde{z}_{t-1} + \Delta x_t \tag{5}$$

initialized at $\tilde{z}_0 = 0$. In particular, we use $C_z = -I_r$ and $\beta = 0.95$.

This choice of β follows from the size and power properties of the subsequently derived Wald test.[3] Extensive Monte Carlo simulations presented in the Online Appendix show that the finite-sample size of the test is very close to the nominal 5% level regardless of the value of β. This holds true for all cases of regressor persistence considered. With respect to the power of the test, we find that it increases monotonically as β increases for all cases considered. This property is also suggested by the $n^{(1+\beta)/2}$ rate of convergence of the IVX estimator in Theorem

[3] To be precise, we follow the convention in prior literature and use the term "size" throughout this study to indicate the "probability of a Type I error" for the various test statistics considered.

A(i) provided in the Appendix. A closer inspection of the reported power plots suggests that starting from low or moderate values of β, there are considerable power gains when we further increase β towards its upper boundary, especially when the true value of A is closer to the null. Given this evidence, we confidently argue that high values of β yield the highest level of power for the Wald test and, at the same time, yield size very close to the nominal 5% level. Therefore, in the empirical implementation of our testing procedure, we use $\beta = 0.95$, which is among the highest values that β can take. Moreover, we strongly advise against using values of β less than 0.9, as they may lead to unnecessary loss of power for the test statistic.[4]

As it is standard in the literature, we assume that the innovations ε_t of the predictive equation (1) are uncorrelated, while allowing for correlation in the innovations of the predictor sequence u_t. The dependence structure of the innovations is formally presented below: part (i) provides assumptions under conditional homoskedasticity; part (ii) accommodates a general form of conditional heteroskedasticity under additional assumptions.

Assumption INNOV. *Let $\epsilon_t = (\varepsilon_t', e_t')'$, with ε_t as in (1), denote an $\mathbb{R}^{m+r}$-valued martingale difference sequence with respect to the natural filtration $\mathcal{F}_t = \sigma(\epsilon_t, \epsilon_{t-1}, ...)$ satisfying*

$$E_{\mathcal{F}_{t-1}}\left(\epsilon_t \epsilon_t'\right) = \Sigma_t \quad a.s. \quad and \quad \sup_{t\in\mathbb{Z}} E\left\|\epsilon_t\right\|^{2s} < \infty \tag{6}$$

for some $s > 1$, where Σ_t is a positive definite matrix. Let u_t in (2) be a stationary linear process

$$u_t = \sum_{j=0}^{\infty} C_j e_{t-j} \tag{7}$$

where $(C_j)_{j\geq 0}$ a sequence of constant matrices such that $\sum_{j=0}^{\infty} C_j$ has full rank and $C_0 = I_r$. We maintain one of the following assumptions:

(i) *$\Sigma_t = \Sigma_\epsilon$ for all t and $\sum_{j=0}^{\infty} \|C_j\| < \infty$.*

(ii) *The process $(\epsilon_t)_{t\in\mathbb{Z}}$ is strictly stationary ergodic satisfying (6) with $s = 2$ and*

$$\lim_{m\to\infty} \left\|Cov\left[\mathrm{vec}\left(\epsilon_m \epsilon_m'\right), \mathrm{vec}\left(\epsilon_0 \epsilon_0'\right)\right]\right\| = 0. \tag{8}$$

[4] We would like to thank an anonymous referee for suggesting this clarification.

The sequence $(C_j)_{j\geq 0}$ *in (7) satisfies*

$$\sum_{j=0}^{\infty} j \left\| C_j \right\| < \infty. \tag{9}$$

The sequence $(\varepsilon_t)_{t\in\mathbb{Z}}$ *admits the following vec-GARCH*(p,q) *representation:*

$$\varepsilon_t = H_t^{1/2}\eta_t, \quad \text{vec}\left(H_t\right) = \overline{\varphi} + \sum_{i=1}^{q} A_i \text{vech}\left(\varepsilon_{t-i}\varepsilon_{t-i}'\right) + \sum_{k=1}^{p} B_k \text{vech}\left(H_{t-k}\right) \tag{10}$$

where $(\eta_t)_{t\in\mathbb{Z}}$ *is an i.i.d.* $(0, I_m)$ *sequence,* $\overline{\varphi}$ *is a constant vector,* A_i*,* B_k *are symmetric positive semidefinite matrices for all* i, k*, and the spectral radius of the matrix* $\Gamma = \sum_{i=1}^{q} A_i + \sum_{k=1}^{p} B_k$ *satisfies* $\rho\left(\Gamma\right) < 1$*.*

Assumption INNOV(i) imposes conditional homoskedasticity on the martingale difference sequence ϵ_t and short-memory on the linear process (7). Assumption INNOV(ii) accounts for conditionally heteroskedastic ϵ_t with finite fourth-order moments of a very general form: the vec-GARCH process in (10) is the most general multivariate GARCH specification (see Chapter 11 of Francq and Zakoian (2010)).[5]

Following standard notational convention, we define the short-run and long-run covariance matrices associated with the innovations ε_t and u_t in (1), (2) as follows:

$$\Sigma_{\varepsilon\varepsilon} = E\left(\varepsilon_t\varepsilon_t'\right), \; \Sigma_{\varepsilon u} = E\left(\varepsilon_t u_t'\right), \; \Sigma_{uu} = E\left(u_t u_t'\right) \tag{11}$$

$$\Omega_{uu} = \sum_{h=-\infty}^{\infty} E\left(u_t u_{t-h}'\right), \; \Omega_{\varepsilon u} = \Sigma_{\varepsilon u} + \Lambda_{u\varepsilon}', \; \Lambda_{u\varepsilon} = \sum_{h=1}^{\infty} E\left(u_t \varepsilon_{t-h}'\right). \tag{12}$$

Note that $\Omega_{\varepsilon u}$ is only a one-sided long-run covariance matrix because ε_t is an uncorrelated sequence by Assumption INNOV. For the same reason, the long-run covariance of the ε_t sequence is equal to the short-run covariance $\Sigma_{\varepsilon\varepsilon}$. Denoting by $\hat{\varepsilon}_t$ the OLS residuals from (1) and by $\hat{u}_t$

[5] The positive semidefinite condition on the matrices A_i, B_k of (10) and the condition on the spectral radius of their sum are part of the standard Boussama (2006) stationarity conditions for the vec-GARCH process; see Theorem 11.5 of Francq and Zakoian (2010). Condition (8) is a mild weak dependence requirement on the process $\text{vec}(\epsilon_t\epsilon_t')$: it is satisfied if ϵ_t is given a vec-GARCH specification analogous to ε_t, but the results in this paper do not require a parametric specification of the conditional variance structure of the e_t process. A general discussion of the rate of decay of the autocovariance function in (8) in the case of univariate conditionally heteroskedastic time series admiting a stationary ARCH(∞) representation is included in Giraitis, Kokoszka and Leipus (2000). The summability condition (9) is standard in the literature on short-memory linear processes (see Phillips and Solo, 1992).

the OLS residuals from (2), the covariance matrices in (11) can be estimated in a standard way:

$$\hat{\Sigma}_{\varepsilon\varepsilon} = \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t\hat{\varepsilon}_t', \quad \hat{\Sigma}_{\varepsilon u} = \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t\hat{u}_t', \quad \hat{\Sigma}_{uu} = \frac{1}{n}\sum_{t=1}^{n}\hat{u}_t\hat{u}_t'. \tag{13}$$

Accommodating autocorrelation in u_t that takes the general form (7) requires non-parametric estimation of the long-run covariance matrices in (12): letting M_n be a bandwidth parameter satisfying $M_n \to \infty$ and $M_n/\sqrt{n} \to 0$ as $n \to \infty$, we employ the usual Newey-West type estimators

$$\hat{\Lambda}_{uu} = \frac{1}{n}\sum_{h=1}^{M_n}\left(1 - \frac{h}{M_n+1}\right)\sum_{t=h+1}^{n}\hat{u}_t\hat{u}_{t-h}', \quad \hat{\Omega}_{uu} = \hat{\Sigma}_{uu} + \hat{\Lambda}_{uu} + \hat{\Lambda}_{uu}' \tag{14}$$

$$\hat{\Lambda}_{u\varepsilon} = \frac{1}{n}\sum_{h=1}^{M_n}\left(1 - \frac{h}{M_n+1}\right)\sum_{t=h+1}^{n}\hat{u}_t\hat{\varepsilon}_{t-h}', \quad \hat{\Omega}_{\varepsilon u} = \hat{\Sigma}_{\varepsilon u} + \hat{\Lambda}_{u\varepsilon}'. \tag{15}$$

Under the full generality of Assumption INNOV, we provide robust inference for the matrix of coefficients A that is invariant to the predictive variables belonging to classes P(i)-P(iv).

Allowing for the presence of an intercept in the predictive equation (1) requires further development of IVX estimation and testing theory. The first step is to use a standard demeaning transformation of (1) that yields exact invariance of estimation of A to the presence of an intercept. We denote sample averages of variates in the system (1)–(2) by $\bar{y}_n = n^{-1}\sum_{t=1}^{n} y_t$, $\bar{x}_{n-1} = n^{-1}\sum_{t=1}^{n} x_{t-1}$, $\bar{\varepsilon}_n = n^{-1}\sum_{t=1}^{n}\varepsilon_t$, the demeaned variates by $Y_t = y_t - \bar{y}_n$, $X_t = x_t - \bar{x}_{n-1}$ and $\mathcal{E}_t = \varepsilon_t - \bar{\varepsilon}_n$, the resulting (demeaned) regression matrices by $\underline{Y} = (Y_1', ..., Y_n')'$ and $\underline{X} = \left(X_0', ..., X_{n-1}'\right)'$, and the (undemeaned) instrument matrix by $\tilde{Z} = \left(\tilde{z}_0', ..., \tilde{z}_{n-1}'\right)'$. We may obtain invariance to the presence of the intercept μ in the predictive equation by subtracting $\bar{y}_n = \mu + A\bar{x}_{n-1} + \bar{\varepsilon}_n$ from (1) and obtaining the transformed predictive equation

$$Y_t = AX_{t-1} + \mathcal{E}_t. \tag{16}$$

We now proceed with IVX estimation of A from the predictive regression system (16), by considering a two-stage least squares estimator based on the near stationary instruments in (5):

$$\tilde{A}_{IVX} = \underline{Y}'\tilde{Z}\left(\underline{X}'\tilde{Z}\right)^{-1} = \sum_{t=1}^{n}\left(y_t - \bar{y}_n\right)\tilde{z}_{t-1}'\left[\sum_{j=1}^{n}\left(x_j - \bar{x}_{n-1}\right)\tilde{z}_{j-1}'\right]^{-1}. \tag{17}$$

Note that the estimator does not involve a demeaned version of the matrix of instruments, as the IVX estimator in (17) is invariant to demeaning $\tilde{z}_{t-1}$ by $\bar{z}_{n-1}$.

The asymptotic behavior of the normalized and centred IVX estimator in (17) is summarized by Theorem A in the Appendix. It turns out that the varying persistence levels of the predictor process in (2) and the effect of estimating an intercept in the predictive model (1) become manifest only in the limit distribution of the normalized signal matrix $\underline{X}'\tilde{Z}$. After appropriate centering and normalization, the $\underline{Y}'\tilde{Z}$ component of the IVX estimator converges in distribution to a Gaussian variate that is independent of the (possibly) random limit in distribution of the signal matrix. As a result, the IVX estimator in (17) follows a *mixed Gaussian limit distribution* irrespective of the degree of persistence of the predictor variable in (2).

The asymptotic mixed normality property of the IVX procedure implies that linear restrictions on the coefficients A generated by the system of predictive equations (1) can be tested by a standard Wald test based on the IVX estimator for all persistence scenarios conforming to the classes P(i)-P(iv). In particular, we consider a set of linear restrictions

$$H_0 : H\text{vec}\,(A) = h, \tag{18}$$

where H is a known $q \times mr$ matrix with rank q and h is a known vector. We propose the following IVX-Wald statistic for testing H_0 in (18):

$$W_{IVX} = \left(H\text{vec}\tilde{A}_{IVX} - h\right)' Q_H^{-1} \left(H\text{vec}\tilde{A}_{IVX} - h\right) \tag{19}$$

where $\tilde{A}_{IVX}$ is the IVX estimator in (17),

$$\begin{aligned} Q_H &= H\left[\left(\tilde{Z}'\underline{X}\right)^{-1} \otimes I_m\right] \mathbb{M} \left[\left(\underline{X}'\tilde{Z}\right)^{-1} \otimes I_m\right] H' \\ \mathbb{M} &= \tilde{Z}'\tilde{Z} \otimes \hat{\Sigma}_{\varepsilon\varepsilon} - n\bar{z}_{n-1}\bar{z}_{n-1}' \otimes \hat{\Omega}_{FM} \qquad (20) \\ \hat{\Omega}_{FM} &= \hat{\Sigma}_{\varepsilon\varepsilon} - \hat{\Omega}_{\varepsilon u}\hat{\Omega}_{uu}^{-1}\hat{\Omega}_{\varepsilon u}' \qquad (21) \end{aligned}$$

and the matrices $\hat{\Sigma}_{\varepsilon\varepsilon}$, $\hat{\Omega}_{\varepsilon u}$ and $\hat{\Omega}_{uu}$ are defined in (13), (14) and (15).

Theorem 1. *Consider the model (1)–(3) under Assumption INNOV with instruments* $\tilde{z}_t$ *defined by (4) and (5). Then, the Wald statistic in (19) for testing (18) satisfies*

$$W_{IVX} \Rightarrow \chi^2(q) \quad as\ n \to \infty$$

under H_0*, for the following classes of predictor processes* x_t *in (2):*

(i) *P(i)-P(iv) under Assumption INNOV(i)*

(ii) *P(i)-P(iii) under Assumption INNOV(ii).*

The proof of Theorem 1 can be found in the Online Appendix. Theorem 1 establishes the robustness of the IVX-Wald test in (19) to the persistence properties of the predictor process in (2). It shows that the IVX methodology provides a unifying framework of inference in predictive regressions that encompasses the whole range of empirically relevant autoregressive data generating mechanisms, from stationary processes to purely nonstationary random walks.

The only class of predictor variables not covered by Theorem 1 is that of purely stationary autoregressions P(iv) with conditionally heteroskedastic innovations. This is by no means surprising since, in the above case, the IVX-Wald test statistic is asymptotically equivalent to a standard Wald statistic of the form:

$$W_n = \left(H\text{vec}\hat{A}_{OLS} - h\right)' \left[H\left\{\left(\underline{X}'\underline{X}\right)^{-1} \otimes \hat{\Sigma}_{\varepsilon\varepsilon}\right\}H'\right]^{-1} \left(H\text{vec}\hat{A}_{OLS} - h\right)$$

with $\hat{A}_{OLS}$ the usual OLS estimator. It is well known that, even with *a priori* knowledge that x_t is a stationary process, W_n will not have a $\chi^2(q)$ limit distribution when the innovation sequence ε_t in (1) is conditionally heteroskedastic because the asymptotic variance of $n^{-1/2}\sum_{t=1}^{n}(x_{t-1} \otimes \varepsilon_t)$ is given by $\Upsilon = E\left(x_{t-1}x'_{t-1} \otimes \varepsilon_t\varepsilon'_t\right)$ and does not factorise to $E\left(x_{t-1}x'_{t-1}\right) \otimes \Sigma_{\varepsilon\varepsilon}$ as in the case when ε_t are conditionally homoskedastic (see equation (35) in Theorem A below). Consequently, the matrix $\left(\underline{X}'\underline{X}\right)^{-1} \otimes \hat{\Sigma}_{\varepsilon\varepsilon}$ is no longer a consistent estimator of the asymptotic variance of $\text{vec}\left(\hat{A}_{OLS}\right)$ and W_n will fail to be asymptotically $\chi^2(q)$.

This failure is a characteristic of least squares regression rather than IVX methodology. It can be rectified by introducing a White (1980)-type of correction in the Wald statistic. In particular, using $\hat{\Upsilon}_n = n^{-1}\sum_{t=1}^{n}\left(\tilde{z}_{t-1}\tilde{z}'_{t-1} \otimes \hat{\varepsilon}_t\hat{\varepsilon}'_t\right)$ as estimator for Υ, with $\hat{\varepsilon}_t$ being the OLS residuals from (1), and replacing the matrix $\mathbb{M}$ in (20) by $\tilde{\mathbb{M}} = n\hat{\Upsilon}_n - n\bar{z}_{n-1}\bar{z}'_{n-1} \otimes \hat{\Omega}_{FM}$ makes the IVX-Wald statistic in (19) heteroskedasticity-robust even in the P(iv) case.

The robustness of the IVX-Wald statistic to conditional heteroskedasticity for the persistence classes P(i)-P(iii) is a novel result of considerable interest: it depends on establishing the invariance, under Assumptions INNOV(i) and INNOV(ii), of the limit distribution in the central limit theorem for $n^{-(1+\alpha)/2} \sum_{t=1}^{n} (x_{t-1} \otimes \varepsilon_t)$ for $\alpha \in (0, 1)$; see Lemma B4 of the Online Appendix. Since the IVX instrument $\tilde{z}_t$ behaves asymptotically as a near stationary process (z_t if $\beta < \alpha$ and x_t if $\beta > \alpha$), Lemma B4 ensures that $\text{vec}\left(\tilde{A}_{IVX}\right)$ will have the same asymptotic variance under Assumptions INNOV(i) and INNOV(ii) for all $\alpha > 0$. The methods developed in Lemma B4 can be used in a wider context to show that any amount of persistence (even of arbitrarily small order) in the regressor x_t alleviates asymptotically the effect of conditional heteroskedasticity and results to t and Wald statistics with standard limit distributions. This phenomenon becomes manifest in long-horizon predictive regressions when the horizon parameter tends to infinity with the sample size; see Theorem 2(ii) in Section 5 below.

Removing the finite-sample distortion to the mixed normal limit distribution of the IVX estimator caused by the estimation of the intercept is another subtle issue. The component Q_H in the quadratic form of the Wald statistic in (19) contains a finite-sample correction in the form of a weighted demeaning of the dominating term $\tilde{Z}'\tilde{Z} \otimes \hat{\Sigma}_{\varepsilon\varepsilon}$ of $\mathbb{M}$ in (20) by $n\bar{z}_{n-1}\bar{z}'_{n-1} \otimes \hat{\Omega}_{FM}$. Despite not contributing to the first-order limit theory for W_{IVX}, this correction removes the finite-sample effects of estimating an intercept in (1). As discussed in Remark A(b) of the Appendix, these effects are more prominent for highly persistent regressors that are strongly correlated with the predictive model's innovations ε_t. Weighting the demeaning in (20) by the long-run covariance matrix $\hat{\Omega}_{FM}$ (that appears in the Phillips and Hansen (1990) FM-endogeneity correction for integrated systems) controls the effect of correlation between ε_t and u_t on the remainder term of the Gaussian first-order approximation (see equation (36) in the Appendix) by the degree of demeaning of the instrument moment matrix $\tilde{Z}'\tilde{Z}$. To obtain better intuition on the nature of the correction in (20), assume for simplicity that $m = r = 1$; then

$$\mathbb{M} = \left[\sum_{t=1}^{n} \tilde{z}_{t-1}^2 - n\bar{z}_{n-1}^2 \left(1 - \hat{\rho}_{\varepsilon u}^2\right)\right] \hat{\Sigma}_{\varepsilon\varepsilon}$$

where $\hat{\rho}_{\varepsilon u} = \hat{\Omega}_{\varepsilon u}/\sqrt{\hat{\Sigma}_{\varepsilon\varepsilon}\hat{\Omega}_{uu}}$ is an estimator of the long-run correlation coefficient between ε_t and u_t. Therefore, the correction in (20) applies a weighting of the demeaning of the $\tilde{Z}'\tilde{Z}$ matrix according to the magnitude of the absolute value of the long-run correlation coefficient $\rho_{\varepsilon u}$, with higher values of $\rho_{\varepsilon u}$ associated with reduced degree of demeaning.

2. Finite-sample properties

2.1 Univariate case

This Section analyzes the finite-sample performance of the IVX-Wald statistic in (19) by means of an extensive Monte Carlo study and compares it to the performance of the Q-statistic of CY and the $\pi^*_{0.05}$ statistic of Jansson and Moreira (2006, henceforth JM). We run two-sided tests with nominal size 5% for all three statistics under the null hypothesis that the slope coefficient in the predictive regression is zero, i.e., $H_0 : A = 0$.

We use the following data generating process (DGP) for the univariate case, where y_t and x_t are scalars. For $t \in \{1, ..., n\}$, the innovation sequences $\varepsilon_t \sim NID(0, 1)$ and $e_t \sim NID(0, 1)$ generate the model:

$$y_t = \mu + Ax_{t-1} + \varepsilon_t \tag{22}$$

$$x_t = R_n x_{t-1} + u_t, \quad R_n = 1 + C/n \tag{23}$$

$$u_t = \phi u_{t-1} + e_t \tag{24}$$

We denote by $\delta = E(\varepsilon_t u_t)$ the contemporaneous correlation coefficient between ε_t and u_t. Simulation results using 10,000 repetitions are presented for values of $C \in \{0, -5, -10, -15, -20, -50\}$, $\delta \in \{-0.95, -0.5, 0, 0.5, 0.95\}$, sample size $n \in \{100, 250, 500, 1000\}$ and $\phi \in \{0, 0.5\}$. In the Online Appendix, we also present simulation results for $\phi = 0.25$ and $\phi = -0.1$, while additional results for $\delta \in \{-0.75, -0.25, 0.25, 0.75\}$ are available upon request. The system is initialized at $x_0 = 0$. The IVX estimator and the corresponding Wald statistic are invariant to the value of μ, so we opt for $\mu = 0$. We consider the following sequence of local alternatives for power comparisons:

$$A = \frac{b}{n}\sqrt{1 - \delta^2} \text{ for } b \in \{0, 2, 4, .., 32, 40, 60, 100\} \tag{25}$$

with $b = 0 \Rightarrow A = 0$ corresponding to the size of each test.

The results regarding the empirical size in the case of no autocorrelation in the predictor's innovation sequence u_t (i.e., $\phi = 0$) are presented in Table 1. We observe that for sample sizes $n \geq 250$ the Wald statistic has excellent size control across all values of C and δ. For $n = 100$, it only appears to be slightly oversized when $|\delta| = 0.95$ and $C \in \{0, -5\}$. For the other combinations of C and δ, the Wald statistic has the correct size. On the other hand, the Q-statistic appears to be undersized for moderate to high values of δ, such as $|\delta| = 0.5$; increasing the sample size does not seem to remedy this problem. Moreover, for autoregressive roots away from unity and very high values of $|\delta|$, the Q-statistic becomes severely oversized; see e.g., the combinations $R_n = 0.5$ and $|\delta| \in \{0.95, 0.5\}$ as well as $R_n = 0.8$ and $|\delta| = 0.95$. This is a manifestation of the fact that the CY procedure is not valid for predictors with low degree of persistence. Finally, the JM statistic also exhibits severe size distortions. The most striking finding is that it becomes extremely oversized across all degrees of regressor persistence when $|\delta| = 0$. In addition, considering high values of $|\delta|$, such as $|\delta| = 0.95$, and autoregressive roots away from unity, the JM statistic becomes severely oversized too. Its size distortions appear to be minimized when $|\delta| = 0.5$.

−Table 1 here−

We subsequently examine the finite-sample size of these three statistics in the presence of autocorrelation in u_t. Table 2 refers to the case where u_t is an AR(1) process with root $\phi = 0.5$. We find that the Wald statistic exhibits size very close to the nominal 5% apart from some slight oversizing for $n = 100$, $C \in \{0, -5\}$ and $|\delta| = 0.95$. On the other hand, the Q-statistic has size substantially lower than the nominal 5% for $|\delta| = 0.5$. Moreover, for $C = -50$, $|\delta| = 0.95$ and $n = 100, 250, 500$, the Q-statistic is severely oversized. Regarding the JM statistic, its size distortions are exacerbated in the presence of autocorrelation. The statistic is severely oversized for both high and low values of $|\delta|$ across all degrees of regressor persistence. As in the case of no autocorrelation, the size distortions of this statistic appear to be minimized when $|\delta| = 0.5$.

−Table 2 here−

We next examine the power properties of the three statistics. Our simulation study computes power with respect to the local alternatives given in (25) without size adjustment, as there is no oversizing in the proposed Wald statistic. We present here results for sample size $n = 250$ and correlations $\delta \in \{-0.95, -0.5, 0\}$, while the corresponding results for $n = 1000$ are presented in the Online Appendix.[6] The power plots presented here correspond to the case of no autocorrelation in u_t (i.e., $\phi = 0$), while in the Online Appendix we present the corresponding power plots for $\phi = 0.5$.

Figure 1 presents the power of the three statistics for $n = 250$, $\delta = -0.95$ and for all values of C considered. We observe that in the unit root case ($C = 0$), the Wald statistic has higher power than the Q-statistic for alternatives close to the null hypothesis but this relationship is reversed for alternatives farther away from the null. For all of the other persistence scenarios (i.e., for all values of $C < 0$), the Wald statistic dominates the Q-statistic for any choice of local alternative and δ. The distance between the power curves of the two statistics increases in favor of the Wald test as the persistence of the regressor is reduced towards stationarity (i.e., as $|C|$ increases). The last Panel of Figure 1 for $C = -50$ shows that despite being considerably oversized in this case, the Q-statistic appears to have lower power in comparison to the (correctly sized) Wald test. Moreover, the JM statistic is characterized by a remarkable lack of power, with the exception of the unit root case. For lower degrees of regressor persistence, the power of the JM statistic is approximately equal to its size even for alternatives far away from the null, undermining further its suitability for predictability tests.

–Figure 1 here–

Figure 2 presents power comparisons for $\delta = -0.5$ and $n = 250$. The power of the Wald test uniformly dominates that of the Q-statistic for all persistence scenarios, including that of a unit root regressor ($C = 0$). As before, the dominance of the IVX over the CY procedure increases as the degree of persistence is reduced towards stationarity. In addition, the power of the JM statistic is much lower relative to the other two statistics, especially as we move away from the unit root case.

–Figure 2 here–

[6] In addition, simulation results for $n \in \{100, 500\}$ and $\delta \in \{-0.75, -0.25\}$ are available upon request. The relative performance of the IVX and CY procedures is very similar to the results reported in this section; the Wald statistic dominates the Q-statistic in terms of power, with the exception of the combinations $\delta \in \{-0.95, 0\}$ and $C = 0$, where there is no dominating relationship.

Figure 3 presents power comparisons for $\delta = 0$ and $n = 250$. The Q-statistic appears to have higher power relative to the Wald statistic in the unit root case ($C = 0$). However, as the degree of persistence is reduced ($C < 0$), the power of the Wald statistic becomes indistinguishable from the power of the Q-statistic. The lack of power for the JM statistic relative to the Wald and the Q-statistic is evident in this case too. Interestingly, this conclusion holds true even in the cases where the JM statistic is severely oversized.

–Figure 3 here–

2.2 Conditionally Heteroskedastic DGP

Recalling that the asymptotic results for the proposed Wald statistic are also valid under conditional heteroskedasticity, we employ a GARCH(1,1) DGP to examine the finite-sample properties of the statistic and compare them with the corresponding properties of the Q-statistic of CY (see the Online Appendix for the DGP specification).

Extensive simulation results are reported in the Online Appendix. We find that the Wald statistic exhibits no size distortion for every parameter combination considered. The Q-statistic exhibits correct size for $\delta = 0$, but it is oversized for the combination $n = 100$, $|\delta| = 0.95$ and $C = -50$, while it is undersized when $|\delta| = 0.5$. With respect to the power of the tests, we find that for $\delta = -0.95$ the Wald statistic dominates the Q-statistic for every degree of regressor persistence considered. The same conclusion is derived for $\delta = -0.5$. For $\delta = 0$, we find that in the unit root case ($C = 0$), the Q-statistic has higher power than the Wald statistic, while for all other degrees of regressor persistence ($C < 0$), the two statistics appear to have the same power.

2.3 Additional Monte Carlo results

We additionally examine the robustness of the power properties of the IVX-Wald statistic with respect to the choice of kernel for the estimation of the long-run covariance matrix. In particular, apart from the Bartlett kernel that we use in the benchmark results, we alternatively use: *i*) the Parzen kernel and *ii*) the Quadratic Spectral kernel. Results are reported in the Online Appendix. In most of the cases, we find that the power plots are almost indistinguishable across the three kernels used.

Finally, we examine the robustness of the power properties of the IVX-Wald statistic when alternative lag lengths are used for the Newey-West estimator of the long-run covariance matrix. In particular, apart from the truncation lag $n^{1/3}$ that we use in the benchmark results, we alternatively consider the following truncation lags: *i)* $n^{1/4}$ and *ii)* $n^{1/2}$, where n is the sample size. Overall, the results presented in the Online Appendix show that the choice of truncation lag yields no considerable difference in terms of power.

2.4 Multivariate case

In this Section, we examine the finite-sample performance of the Wald statistic in the context of multivariate regressions. We generalise the DGP of Section 2.1 to include more than one predictors. In particular, we use the following DGP:

$$y_t = \mu + Ax_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim NID(0,1), \tag{26}$$

$$x_t = R_n x_{t-1} + u_t, \quad R_n = I_r + C/n, \tag{27}$$

$$u_t = \Phi u_{t-1} + e_t, \quad \Phi = diag\left(\phi_1, \phi_2, \phi_3\right), \quad e_t \sim N\left(0, \Theta\right), \tag{28}$$

$$\Sigma = E\left(\zeta_t \zeta_t'\right), \quad \zeta_t = \left(\varepsilon_t, u_t'\right)' \tag{29}$$

where x_t is a 3×1 vector that contains three regressors. Each regressor is characterised by a different degree of persistence. In particular, we set $C = diag\,(0, -10, -100)$, corresponding to a unit root, a local-to-unity and a stationary regressor.[7]

To render the examined setup empirically relevant, we use values for Φ and Σ that are estimated from a predictive system with the CRSP S&P 500 log excess returns being the regressand and the earnings-price ratio (unit root), T-bill rate (local-to-unity) and inflation rate (stationary) being the regressors. In particular, Correlation Set 1 corresponds to the correlation structure of the residuals (δ's) and autocorrelation coefficients (ϕ's) that are estimated from monthly data during the full sample period, while the corresponding parameters of Correlation Set 2 are estimated from quarterly data. In addition to these parameters, we also examine the size properties of the Wald test when alternatively $\Phi = 0_{3\times 3}$ (i.e., no autocorrelation), $\Phi = 0.25I_3$ and $\Phi = 0.5I_3$. Finally, we consider sample sizes $n \in \{250, 500, 1000\}$.

[7] We would like to thank an anonymous referee for suggesting this setup.

We examine the size properties of four different tests using a 5% significance level. The first one is the joint Wald test (W_{joint}) under the null hypothesis that all three slope coefficients are jointly equal to zero, i.e., $H_0 : A = (0, 0, 0)$ in (26). The other three tests refer to the individual significance of each regressor. In particular, W_{UR} corresponds to the Wald test under the null hypothesis that the slope coefficient of the unit root regressor is equal to zero, i.e., $H_0 : A_1 = 0$, letting the other two slope coefficients free. Similarly, W_{LTU} corresponds to the Wald test under $H_0 : A_2 = 0$ and $W_{Stationary}$ corresponds to the Wald test under $H_0 : A_3 = 0$.

Table 3 presents the finite-sample size of these four Wald tests. For Correlation Set 1 in the upper Panel, we find that the size of the joint Wald test (W_{joint}) is very close to the nominal 5% across all autocorrelation structures examined. With respect to the test of individual significance for the unit root regressor (W_{UR}), we find a slight oversizing, which peaks around 8%. However, this oversizing becomes almost negligible for the test of individual significance for the local-to-unity regressor (W_{LTU}), while the corresponding test for the stationary regressor ($W_{Stationary}$) exhibits no size distortion. Examining the size properties using Correlation Set 2 in the lower Panel of Table 3, we find no size distortion across these four tests, regardless of the autocorrelation structure used.

–Table 3 here–

We also examine the power properties of the joint Wald test under the null hypothesis $H_0 : A = (0, 0, 0)$, as the slope coefficient of each of the three regressors increases. In particular, $\text{Wald}_{0.05}^{UR}$ refers to the power of the joint test when, under the alternative, the slope coefficient of the unit root regressor takes non-zero values $\left(A = \frac{b}{n}(1, 0, 0)\right)$, $\text{Wald}_{0.05}^{LTU}$ refers to the power of the joint test when the slope coefficient of the local-to-unity regressor increases $\left(A = \frac{b}{n}(0, 1, 0)\right)$, while $\text{Wald}_{0.05}^{Stationary}$ refers to the power of the joint test when the slope coefficient of the stationary regressor increases $\left(A = \frac{b}{n}(0, 0, 1)\right)$. Local alternatives are derived using $b \in \{0, 2, 4, ..., 32, 40, 60, 100\}$ with $b = 0$ corresponding to the size of the test, while we consider $n \in \{100, 250, 500, 1000\}$.[8]

Figure 4 presents the power plots of the joint Wald statistic using Correlation Set 1, while Figure 5 presents the corresponding power plots using Correlation Set 2. We find that in every

[8]Simulation results for the power properties of the individual significance tests in the presence of multiple regressors are available upon request.

case examined, the joint Wald test has very good power properties, since the rejection rate monotonically increases as the true value of the corresponding slope coefficient increases. This holds true for all sample sizes examined. Moreover, the power of the joint Wald test is remarkably high even for low values of local alternatives for the slope coefficient of the unit root and the local-to-unity regressors.

–Figures 4 and 5 here–

3. Data and regressors' degree of persistence

We implement the proposed methodology to test the predictive ability of commonly used financial variables with respect to excess stock market returns. The examined period is 1927–2012. The employed dataset is an updated version of the one used in Welch and Goyal (2008).[9] For our benchmark predictability tests, we use monthly and quarterly data, while in the Online Appendix we report results for annual data too. Following Welch and Goyal (2008), we use S&P 500 value-weighted log excess returns to proxy for excess market returns. Moreover, we use the following 12 variables as potential predictors: T-bill rate **(tbl)**, long-term yield **(lty)**, term spread **(tms)**, default yield spread **(dfy)**, dividend-price ratio **(d/p)**, dividend yield **(d/y)**, earnings-price ratio **(e/p)**, dividend payout ratio **(d/e)**, book-to-market value ratio **(b/m)**, net equity expansion **(ntis)**, inflation rate **(inf)** and consumption-wealth ratio **(cay)**. We present the definitions of these variables as well as a list of prior studies that have examined their predictive ability in the Online Appendix. It should be noted that cay is only available at quarterly and annual frequency, starting from 1952 for quarterly and 1945 for annual data.

One of the main advantages of the IVX methodology is that, by virtue of its robustness, it does not require any pre-testing to determine the degree of predictors' persistence prior to conducting predictability tests. Pre-testing procedures naturally increase the Type I error of predictability tests and may well lead to conflicting empirical conclusions. To demonstrate this point, we report for each regressor in Table 4 the least squares point estimate of the autoregressive root $\widehat{R}_n$ from regression (2) using monthly data as well as the results of four unit root tests that are commonly used as pre-tests: the Augmented Dickey Fuller (ADF) test, the DF-GLS test

[9] The dataset updated up to December 2012 is sourced from Amit Goyal's website: http://www.hec.unil.ch/agoyal/

by Elliott, Rothenberg and Stock (1996), the Phillips-Perron (PP) test as well as the KPSS test by Kwiatkowski, Phillips, Schmidt and Shin (1992); the lag length for ADF and DF-GLS is determined by the Bayesian information criterion. It is remarkable how close to unity the estimated root is for most of the variables: for d/y, d/p and e/p the estimated root is exactly equal to unity up to three decimal points. The four pre−tests agree on the existence of a unit root only for lty, d/y and d/p. For the remaining variables, the tests yield contradictory results. Even for inf, which exhibits a relatively low autoregressive root, the KPSS test would reject the null hypothesis of no unit root at the 5% level.

−Table 4 here−

Table 5 contains the corresponding results for quarterly data, confirming that these variables exhibit a very high degree of persistence, even when they are measured at a lower frequency, and that their autoregressive root is very close to unity, with the exception of inf. Interestingly, cay also exhibits a very high autoregressive root and the ADF and PP tests would not reject the null hypothesis of unit root. The evidence provided in the Online Appendix using annual data is very similar, though the autoregressive coefficients are somewhat lower for some variables. Overall, neither the conclusions of the pre-tests nor the estimated autoregressive roots alleviate the uncertainty on the exact degree of persistence of the employed regressors, regardless of the frequency used. This observation, along with Type I error considerations, motivates further the use of the proposed IVX econometric framework.

−Table 5 here−

4. Predictability tests

4.1 Univariate regressions

4.1.1 Monthly data We firstly examine the individual predictive ability of each of the employed regressors using monthly data. Table 6 contains the results for these univariate regressions using the proposed IVX estimator and the corresponding Wald statistic under the null hypothesis of no predictability. For comparison, we also report: i) the t-ratio under the standard least squares approach, ii) the 90% Bonferroni confidence interval from the Q-statistic of CY and

iii) the p-value for the JM statistic. Moreover, we report the correlation coefficient (δ) of the residuals from regressions (1) and (2) as a measure of the regressor's degree of endogeneity.

−Table 6 here−

Panel A reports the results for the period January 1927–December 2012. Using our test statistic, we find that the null of no predictability can be rejected at the 5% level only when the lagged e/p, b/m and ntis are used as predictors; d/y is significant only at the 10% level. To the contrary, there is no evidence of significant predictive ability for d/e, lty, d/p, tbl, dfy, tms and inf in the full sample period. Comparing our findings with the other test statistics, there are important differences with respect to which predictors are significant and at what level. Standard least squares inference indicates that d/y is significant at the 5% and that ntis is significant only at the 10% level. More interestingly, the Q-test of CY fails to report the significance of e/p even at the 10% level. Calculating 95% Bonferroni confidence intervals for the Q-test in unreported results we find that only ntis is significant at the 5% level. These findings are in line with our simulation results for the size properties of the Q-test, where we documented that it tends to underreject for large sample sizes ($n = 1,000$) and for moderate to high degrees of endogeneity, such as the one estimated for e/p ($\delta = -0.76$). Finally, the JM test does not find e/p or b/m to be significant predictors, while it does so for tbl and dfy, which are insignificant according to our test.

Panel B of Table 6 reports the corresponding results for the period after 1952. This subperiod is examined for two reasons. Firstly, the term structure variables (tbl, tms and lty) are thought to be more informative since the Fed abandoned its policy of pegging the interest rate (1951 Treasury Accord). Moreover, cay becomes available at quarterly frequency during this period. Secondly, prior studies (see e.g., CY) have found that the evidence of predictability has weakened in more recent sample periods, and hence it can be attributed to early periods when such patterns were not documented. The proposed testing methodology can shed further light on this conjecture.

In fact, the predictability evidence almost entirely disappears in the post–1952 period. The IVX-Wald test indicates that only inf is significant at the 5% level. Moreover, tbl and tms are significant at the 10% level, supporting the argument that the term structure variables may have become more informative after 1952. Similar is the evidence based on the Q-test of CY, with the main difference that their test additionally finds d/y to be marginally significant at the 10% level. More striking are the differences with least squares inference, according to which both d/y and d/p are significant at the 10% level, while tbl is significant at the 5% level, demonstrating its tendency to overreject the null of no predictability. Using the JM test would also lead to conclusions that are considerably different from ours. Most importantly, this test indicates d/y as a significant predictor, while it fails to do so for tbl and inf. Overall, our results support the argument that predictability has considerably weakened, if not disappeared, after 1952.

4.1.2 Quarterly data We subsequently estimate the univariate predictive regressions using quarterly data and we report the corresponding results in Table 7 for the full sample period (Panel A) and the post–1952 period (Panel B), respectively. The results are very similar to the ones we derived using monthly data. In particular, the IVX-Wald test indicates that in the full sample period, e/p, b/m and ntis are again found to be significant predictors at the 5% level, while we also report significance for d/p at the 10% level. Standard least squares inference would point to similar conclusions, with even lower p-values due to the tendency of the t-test to overreject. More striking is the comparison with the inference derived from the Q-test. In particular, the latter fails to find significance for either e/p or b/m even at the 10% level, demonstrating again a tendency to underreject for moderate to high values of δ. The inference derived from the JM test is also very different from ours. In particular, the JM test fails to report significance for e/p and d/p, while it indicates d/y and dfy as strongly significant.

–Table 7 here–

For the post–1952 period we find that, according to the IVX-Wald test, only tms out of the previously used variables remains significant at the 10% level. The rest of the tests also show that predictability has overall weakened in this subperiod, but they additionally find some other variables to be significant predictors, at least at the 10% level. The most interesting finding is that cay, which becomes available after 1952, is a highly significant predictor across all tests considered, including our Wald test. This striking finding corroborates the results of Lettau and Ludvigson (2001) for the updated sample period that we examine.

Taken together with the corresponding univariate results for annual data reported in the Online Appendix, the Wald test indicates that there is significant evidence of in-sample predictability for e/p, b/m and ntis in the full sample period, and weaker evidence for the dividend-based ratios. However, this evidence almost entirely disappears during the post–1952 period, with the exception of some rather weak evidence for the term structure variables (tms and tbl). The only variable that is found to be strongly significant in the post–1952 period is cay.

4.2 Multivariate regressions

The previous Section considered univariate predictability tests. However, it is common practice to employ multiple regressors and to assess their joint significance; this approach is informative for market efficiency tests since predictability should be assessed with respect to the entire information set, not each variable in isolation (see also Cochrane, 2011, for a discussion of the multivariate challenge in predictability tests). Moreover, multivariate predictive regressions are widely used in VAR systems for intertemporal asset pricing tests (e.g., Campbell and Vuolteenaho, 2004) as well as in conditional performance evaluation studies (e.g., Ferson, Sarkissian and Simin, 2008). Additionally, from a theoretical point of view, recently developed present value models (see e.g., Ang and Bekaert, 2007, and Golez, 2014) suggest that d/p alone cannot capture the variation in expected stock returns due to stochastic discount rates and/or dividend growth, and hence it should be used jointly with other predictors.

Given the importance of multivariate predictive systems, it is unfortunate that the recent methodological contributions that correct for the bias arising in least squares inference are developed for univariate regressions only. The notable exception is the iterative procedure of Amihud et al. (2009), which is based on the augmented regression method of Amihud and Hurvich (2004) and accommodates multiple regressors in a single-horizon predictive setup under the restriction that the predictors are stationary. Their procedure yields a reduced-bias estimator and the corresponding test statistic is shown to have good size properties, which, however, deteriorate as the persistence of the predictors approaches the nonstationarity boundary.

On the other hand, our instrumental variable approach introduces an easy-to-implement Wald statistic, enabling us to conduct valid inference regardless of the dimensionality of the predictive system and for all known types of regressors' persistence, from strictly stationary to unit root processes, while it is also applicable to long-horizon predictive systems. The proposed Wald test allows us to examine the joint as well as the individual significance of the regressors used in a multivariate system. In particular, to test their joint significance, we compute the Wald statistic (19) under the null hypothesis that all slope coefficients are equal to zero, i.e., $H_0 : A = 0_{1xr}$, while the individual significance of each predictor is evaluated under the null hypothesis that the corresponding slope coefficient is equal to zero, i.e., $H_0 : A_i = 0$.

We utilize this test to re-examine the predictive ability of certain combinations of regressors that were found to be significant in prior studies and they are motivated from either a theoretical or an empirical point of view. In particular, we use the following combinations: i) d/p and tbl (Ang and Bekaert, 2007), ii) d/p, tbl, dfy and tms (Ferson and Schadt, 1996), iii) d/p and b/m (Kothari and Shanken, 1997), iv) d/p and d/e (Lamont, 1998) and v) e/p, tms and b/m (Campbell and Vuolteenaho, 2004). Additionally, we follow a general-to-specific statistical approach to come up with the best set of predictors. In particular, starting with a base system that includes d/p, e/p, tbl, tms, dfy and ntis, we eliminate in each estimation round the variable exhibiting the lowest (and insignificant) value of individual Wald statistic. This process is repeated until all remaining variables are individually significant at the 10% level or lower.[10]

4.2.1 Monthly data Table 8 reports the results for monthly data. Panel A contains the results for the full sample period. Interestingly, we find that none of the examined combinations leads to joint significance at the 5% level. Only the combination of e/p, b/m and tms is jointly significant at the 10% level, but none of these predictors' coefficients is individually significant. It is also noteworthy that d/p is individually insignificant in all combinations examined, apart from the case where it is used jointly with d/e. This finding casts more doubt on its predictive ability over short-horizon returns. On the other hand, the general-to-specific approach leads to a rather interesting finding: e/p and tbl are both jointly and individually significant at the 5% level.

−Table 8 here−

Panel B reports the corresponding results for the post–1952 period, leading to very similar conclusions. None of the five combinations considered is found to be jointly significant and d/p is individually insignificant in every case. Only tbl and tms are found to be individually significant in some cases, confirming that term structure variables may be indeed more informative in the

[10] We would like to thank the Editor for suggesting this approach.

post–1952 period. The general-to-specific approach yields again the most interesting result: e/p and tbl are jointly and individually significant during this subperiod too. As a robustness test, we have alternatively included b/m instead of d/p in the base system; unreported results show that we still end up with e/p and tbl being the only two individually and jointly significant predictors in both periods.

4.2.2 Quarterly data We repeat the previous analysis using quarterly data and we report these results in Table 9. Panel A contains the full sample period results. We find that combinations that include b/m lead to joint significance, but the regressors' coefficients are insignificant. Moreover, we find that d/p is individually significant in some combinations, but none of these yields joint significance. On the other hand, according to the general-to-specific approach, e/p, tbl and ntis are both individually and jointly significant.

–Table 9 here–

Panel B reports the corresponding results for the post–1952 period. Interestingly, we find that none of these five combinations yields joint significance. Since cay becomes available in this subperiod, we additionally examine the combination of d/p, d/e and cay, which was considered in Lettau and Ludvigson (2001). In fact, we find that this combination and cay's coefficient are significant at the 1% level. Moreover, we also include cay in the base system for the general-to-specific approach, given its strong significance in univariate tests. This approach yields a highly significant combination of e/p, tbl, cay and dfy for this subperiod.

Taken together, the multivariate results for monthly and quarterly data show that commonly used combinations of these regressors have limited predictive ability, especially in the post–1952 period. However, a general-to-specific approach indicates that the combination of e/p and tbl is highly significant and robust to the choice of data frequency and the examined period. Finally, these results confirm that cay is a highly significant predictor in the post–1952 period and this significance is not subsumed by other commonly used variables.

5. Long-horizon predictive regressions

The previous tests examined short-horizon predictability using 1-period ahead returns. A related debate in the literature refers to the existence of long-horizon predictability. In particular, a number of studies have found that the predictive ability of certain financial variables becomes stronger as the horizon increases (see inter alia the surveys of Cochrane, 1999, and Campbell, 2000). On the other hand, some recent studies cast doubt on this prevailing view (see Valkanov, 2003, Torous et al., 2004, Ang and Bekaert, 2007, Boudoukh et al., 2008 and Hjalmarsson, 2011). In particular, Ang and Bekaert (2007) find no evidence of long-horizon predictability using standard errors based on the reverse regression approach of Hodrick (1992), which removes the moving average structure in the error term induced by summing returns over long horizons, and hence retains the correct size, as compared to Hansen-Hodrick (1980) and Newey-West (1987) standard errors that lead to severely oversized test statistics.[11] Moreover, Valkanov (2003) and Boudoukh et al. (2008) show that in the presence of highly persistent regressors, predictability may artificially emerge in standard least squares regressions as the horizon increases. We contribute to this debate by extending the proposed IVX-Wald test to accommodate long-horizon predictive regressions and conducting the corresponding empirical tests.[12] Section 5.1 develops a long-horizon IVX-Wald test, Section 5.2 examines the finite-sample properties of the newly developed Wald test, while Section 5.3 discusses the corresponding empirical results.

5.1 Long-horizon IVX inference

Long-horizon predictability tests are typically based on estimators derived from regressing the K-period cumulative stock return $y_t(K)$ on a lagged predictor x_{t-1} and an intercept as in the following fitted model:

$$y_t(K) = \mu_f + Ax_{t-1} + \eta_{f,t} \quad t \in \{1, ..., n - K + 1\} \tag{30}$$

with $y_t(K) = \sum_{i=0}^{K-1} y_{t+i}$, while the DGP characterizing the true relationship between y_t and x_t continues to be given by (1). For brevity, we introduce the notation $v_t(K) = \sum_{i=0}^{K-1} v_{t+i}$ for any sequence $(v_t)_{t\geq 1}$ and let $n_K = n - K + 1$.

It is clear that the accumulation of predicted variables on the left side of (30) generates additional correlations that are not present in short-horizon regressions and affect the stochastic properties of long-horizon estimators. To fix ideas, assume temporarily that the intercepts μ

[11] The recent study of Wei and Wright (2013) extends the reverse regression approach of Hodrick (1992) to a wider range of null hypotheses even when the predictors are local-to-unity processes.

[12] We would like to thank the Editor for suggesting the extension of IVX methodology to the long-horizon case.

in (1) and μ_f in (30) are equal to zero. Then, the OLS estimator of A from (30) is given by $\hat{A}_{OLS}(K) = \sum_{t=1}^{n_K} y_t(K) x'_{t-1} \left(\sum_{t=1}^{n_K} x_{t-1} x'_{t-1}\right)^{-1}$. Using the DGP (1), it is easy to see that the above OLS estimator is inconsistent for $K > 1$:

$$\hat{A}_{OLS}(K) = \left[A \sum_{t=1}^{n_K} x_{t-1}(K) x'_{t-1} + \sum_{t=1}^{n_K} \varepsilon_t(K) x'_{t-1}\right] \left(\sum_{t=1}^{n_K} x_{t-1} x'_{t-1}\right)^{-1},$$

the inconsistency occurring because $\sum_{t=1}^{n_K} x_{t-1}(K) x'_{t-1}$ has the same order of magnitude as $\sum_{t=1}^{n_K} x_{t-1} x'_{t-1}$ for fixed horizon K and dominates $\sum_{t=1}^{n_K} x_{t-1} x'_{t-1}$ asymptotically when $K \to \infty$. This imbalance can be easily corrected by modifying the OLS estimator:

$$\hat{A}_{mOLS}(K) = \sum_{t=1}^{n_K} y_t(K) x'_{t-1} \left(\sum_{t=1}^{n_K} x_{t-1}(K) x'_{t-1}\right)^{-1}. \tag{31}$$

While this modified OLS estimator is consistent, the limit distribution of $\hat{A}_{mOLS}(K) - A$ (under suitable normalisation) will not be mixed Gaussian in the case of unit root and local-to-unity regressors. Consequently, inference procedures based on $\hat{A}_{mOLS}(K)$ will not be valid across the range of persistence classes P(i)-P(iv) of Section 1, leading to erroneous empirical conclusions in the case of misspecification of regressor persistence. IVX methodology can be adapted to deliver robust inference in long-horizon predictive regression systems. The key idea is the same as in the short-horizon case: given a consistent least squares estimator, the IVX estimator is constructed as a feasible instrumental variables estimator that replaces the regressor x_t by the IVX instrument $\tilde{z}_t$ in (31):

$$\hat{A}_{IVX}(K) = \sum_{t=1}^{n_K} y_t(K) \tilde{z}'_{t-1} \left(\sum_{t=1}^{n_K} x_{t-1}(K) \tilde{z}'_{t-1}\right)^{-1}.$$

In the general case where the intercept terms μ in (1) and μ_f in (30) are non-zero, a standard result on partitioned regression yields that least squares estimation of A from the regression (30) is equivalent to least squares estimation of A from the regression:

$$y_t(K) - \bar{y}_{n_K}(K) = A\left(x_{t-1} - \bar{x}_{n_K-1}\right) + \vartheta_t \quad t \in \{1, ..., n_K\} \tag{32}$$

where $\bar{y}_{n_K}(K) = n_K^{-1} \sum_{t=1}^{n_K} y_t(K)$ and $\bar{x}_{n_K-1}(K) = n_K^{-1} \sum_{t=1}^{n_K} x_{t-1}(K)$ denote the sample means of $y_t(K)$ and $x_{t-1}(K)$ and $\bar{y}_{n_K}$ and $\bar{x}_{n_K-1}$ denote the usual sample means of y_t and x_{t-1} based on the first n_K observations, respectively. We define the data matrices $\underline{X}_{n_K-1} = \left[x_0' - \bar{x}_{n_K-1}', ..., x_{n_K-1}' - \bar{x}_{n_K-1}'\right]'$, $\tilde{Z}(K) = \left[\tilde{z}_0'(K), ..., \tilde{z}_{n_K-1}'(K)\right]'$,

$$\begin{aligned} \underline{Y}(K) &= \left[y_1'(K) - \bar{y}_{n_K}'(K), ..., y_{n_K}'(K) - \bar{y}_{n_K}'(K)\right]' \\ \underline{X}(K) &= \left[x_0'(K) - \bar{x}_{n_K-1}'(K), ..., x_{n_K-1}'(K) - \bar{x}_{n_K-1}'(K)\right]' \end{aligned}$$

and $\tilde{Z}_{n_K-1} = \left[\tilde{z}_0', ..., \tilde{z}_{n_K-1}'\right]'$, where, as before, underlying indicates demeaning. The modified OLS estimator from (30) (equivalently from (32)) can be expressed as:

$$\tilde{A}_{mOLS}(K) = \underline{Y}(K)' \underline{X}_{n_K-1} \left[\underline{X}(K)' \underline{X}_{n_K-1}\right]^{-1}$$

and the corresponding IVX estimator of A is given by:

$$\tilde{A}_{IVX}(K) = \underline{Y}(K)' \tilde{Z}_{n_K-1} \left[\underline{X}(K)' \tilde{Z}_{n_K-1}\right]^{-1}. \tag{33}$$

The asymptotic behavior of the normalized and centred IVX estimator in (33) is summarized by Theorem B in the Appendix; asymptotic mixed Gaussianity is preserved irrespective of the degree of persistence of the predictor variable in (2), as long as the rate of growth of the horizon K is slower than that of the sample size n. This requirement is presented formally below.

Assumption H. *The horizon K may be a fixed integer or a sequence $(K_n)_{n\in\mathbb{N}}$ that increases to infinity slower than the sample size n: $K_n/n \to 0$ as $n \to \infty$.*

As in the short-horizon case, the asymptotic mixed normality property of the long-horizon IVX estimator $\tilde{A}_{IVX}(K)$ implies that the associated IVX-Wald test statistic will have a standard chi-squared limit distribution across the whole range of empirically relevant persistence classes P(i)-P(iv). In particular, we propose the following IVX-Wald statistic for testing the set of linear restrictions (18) in long-horizon predictive regression systems:

$$W_{IVX}(K) = \left[H\text{vec}\tilde{A}_{IVX}(K) - h\right]' Q_{H,K}^{-1} \left[H\text{vec}\tilde{A}_{IVX}(K) - h\right] \tag{34}$$

where

$$
\begin{aligned}
Q_{H,K} &= H\left[\left(\tilde{Z}_{n-K}'\underline{X}(K)\right)^{-1}\otimes I_m\right]\mathbb{M}_K\left[\left(\underline{X}(K)'\tilde{Z}_{n-K}\right)^{-1}\otimes I_m\right]H' \\
\mathbb{M}_K &= \tilde{Z}(K)'\tilde{Z}(K)\otimes\hat{\Sigma}_{\varepsilon\varepsilon}-n_K\bar{z}_{n_K-1}(K)\,\bar{z}_{n_K-1}'(K)\otimes\hat{\Omega}_{FM}
\end{aligned}
$$

$\bar{z}_{n_K-1}(K)=n_K^{-1}\sum_{t=1}^{n_K}\tilde{z}_{t-1}(K)$ and $\hat{\Omega}_{FM}$ is defined in (21).

Theorem 2. *Consider the model (1)–(3) under Assumption INNOV with (9) and H. Then, the IVX-Wald statistic in (34) for testing (18) satisfies*

$$
\tilde{W}_{IVX}(K)\Rightarrow\chi^2(q)\quad \text{as } n\to\infty
$$

under H_0 for the following classes of predictor processes x_t in (2):

(i) *P(i)-P(iv) under Assumption INNOV(i)*

(ii) *P(i)-P(iv) under Assumption INNOV(ii) when $K\to\infty$*

(iii) *P(i)-P(iii) under Assumption INNOV(ii) when the horizon parameter K is fixed.*

Theorem 2 shows that, under Assumption H, the robustness property of IVX methodology extends to long-horizon predictive regressions. Note that when $K=1$, the long-horizon IVX estimator (33) and the associated IVX-Wald statistic (34) reduce to their short-horizon counterparts (17) and (19), respectively. Note also the robustness that the IVX-Wald statistic exhibits to conditional heteroskedasticity for purely stationary regressors when $K\to\infty$: this is due to the persistence that the horizon K induces in the predictive regression; see the discussion in the penultimate paragraph of Section 1.

5.2 Finite-sample properties

In this Section, we examine the finite-sample properties of the long-horizon Wald statistic in (34) that corresponds to the long-horizon predictive regression in (30). For this Monte Carlo study, we use the DGP specified in (22)–(24) for the univariate case. In particular, we consider the following parameter values: $C\in\{0,-5,-10,-15,-20,-50\}$, $\delta\in\{-0.95,-0.5,0\}$, $n\in\{100,500,1000\}$ and $\phi=0$. For sample size $n=100$, we consider predictive horizons $K=$

$2, 3, 4, 5$, for $n = 500$ we consider $K = 4, 8, 12, 20$, while for $n = 1000$ we use $K = 4, 12, 36, 60$. Table 10 presents the finite-sample size of the long-horizon Wald statistic. These simulation results show that the size of the proposed test is remarkably close to the nominal 5% level for all cases considered.

−Table 10 here−

We also examine the power properties of the long-horizon Wald statistic, using local alternatives $A = \frac{b}{n}$ with $b \in \{0, 2, 4, .., 32, 40, 60, 100\}$. Power plots for sample size $n = 1000$ and horizons $K = 12, 36, 60$ as well as for sample size $n = 500$ and horizons $K = 4, 12, 20$ are presented in the Online Appendix.[13] In sum, these plots show that for all horizons considered, the power of the statistic rapidly increases as the true value of A increases. Moreover, in each case, the power of the statistic decreases as the predictive horizon increases, but this decrease is very small for highly persistent regressors.

5.3 Empirical results

Table 11 reports the results from long-horizon univariate predictability tests using monthly data. In the full sample period (Panel A), we find no evidence that predictability becomes stronger as the horizon increases, with the exception of tms. To the contrary, the predictive ability of e/p and b/m weakens, being significant only at the 10% level when we examine horizons longer than 12 and 36 months, respectively. Only tms and ntis are significant at the 5% level when we examine a 60-month horizon. Regarding d/y and d/p, these are not significant at the 5% level regardless of the examined horizon.[14] In the post–1952 period (Panel B), predictability almost entirely disappears, especially for horizons beyond 24 months. We find that only d/e becomes significant at long horizons, while tms remains marginally significant at the 10% level.

−Table 11 here−

Table 12 reports the corresponding long-horizon tests using quarterly data. In the full sample period (Panel A), predictability becomes weaker as the horizon increases. Interestingly, e/p, b/m

[13] The corresponding power plots for $n = 100$ are available upon request.

[14] To the contrary, in unreported results we find that using Newey-West or Hansen-Hodrick standard errors to calculate least squares t-ratios, d/y and d/p (as well as most of the other variables) would erroneously appear as highly significant for horizons of 12 months or longer. The findings are similar when we consider quarterly data.

and ntis, which were found to be strongly significant in predicting 1-quarter ahead returns (see Table 7), become less significant as the horizon increases and they are eventually insignificant at the 20-quarter horizon; d/p remains significant at the 10% level for all horizons considered, while tms becomes marginally significant at very long horizons. In the post–1952 period (Panel B), there is no evidence of predictability with three exceptions: tms remains significant but only at the 10% level, d/e becomes marginally significant beyond 8 quarters, while cay is the only variable that remains significant at the 5% level for all horizons examined. Similar is the pattern of the corresponding results using annual data that are reported in the Online Appendix.

–Table 12 here–

In sum, our evidence is in line with the results of the above cited studies that cast doubt on the ability of commonly used variables to predict stock returns at long horizons, especially in the post–1952 period. We actually find that, if anything, predictability is generally weaker, not stronger, as the horizon increases.

Table 13 presents the results for long-horizon predictability tests with multiple regressors. We present only the combinations of regressors that were found to be both individually and jointly significant under the general-to-specific approach described in Section 4.2 and reported in Tables 8 and 9, using 1-month and 1-quarter ahead returns, respectively. Panel A reports the results for monthly data. In the full sample period, we find that while e/p remains individually significant, tbl becomes insignificant as the horizon increases. Their joint predictive ability remains significant, but only at the 10% level for horizons beyond 12 months. For the post–1952 period results are more striking: e/p and tbl are neither individually nor jointly significant beyond 12 months.

–Table 13 here–

Using quarterly data in Panel B, we get a similar pattern. For the full sample period, only e/p remains individually significant for all the examined horizons, while neither tbl nor ntis are significant for longer than 8-quarter horizons; the joint significance of these three variables becomes weaker as we increase the predictive horizon and eventually disappears at the 20-quarter horizon. For the post–1952 period, we find that at horizons longer than 4 quarters, only cay is individually significant at the 5% level, driving the joint significance of the corresponding multivariate system. Overall, our evidence is in broad agreement with the results of Ang and Bekaert (2007), who found that tbl can predict future stock returns within a multivariate setup only at short (less than 1-year) horizons.

6. Conclusion

This study revisits the popular issue of stock return predictability via lagged financial variables. We conduct a battery of predictability tests for US stock returns during the 1927–2012 period, proposing a novel methodology, termed as IVX estimation, which is robust to the time series properties of the employed regressors. The uncertainty regarding the order of integration of these predictive variables has been characterized as a main source of concern for invalid inference, especially in the presence of endogenous regressors (see Stambaugh, 1999, and CY); the robust methodology we propose successfully addresses this concern. In univariate tests, we find that the earnings-price and book-to-market value ratios as well as net equity expansion are significant predictors of 1-period ahead excess market returns. However, this evidence almost entirely disappears in the post–1952 period. Only the consumption-wealth ratio is found to be strongly significant in this subperiod.

Apart from robustifying inference in predictability tests, this novel methodology presents two additional, particularly attractive features. Firstly, it leads to standard chi-squared inference, and hence the construction of Bonferroni-type confidence intervals is avoided. Such a simplification is mostly welcome for practical purposes, given the large number of predictive regressors that have been employed in prior literature. Secondly, the IVX estimation methodology is applicable to multivariate systems of both regressors and regressands. This facility allows us to test a wide range of predictability relationships. Most obviously, we can test for the joint ability of a set of regressors to predict stock market returns. While this issue was the main motivation of the early studies in the literature (e.g., Fama and French, 1989), most of the recently suggested econometric methodologies have been restricted to setups with a scalar regressor (see Torous et al., 2004, CY, JM, and Hjalmarsson, 2011). Our multivariate tests document that the combination of the earnings-price ratio and T-bill rate is highly significant and robust to the choice of data frequency and examined period.

Interestingly, the proposed testing procedure can be extended to long-horizon predictive regressions. We develop the relevant test statistic and we show that it exhibits very good finite-sample properties. Using this newly developed statistic, our long-horizon tests document that, if anything, predictability becomes weaker, not stronger, as the horizon increases. Only the consumption-wealth ratio remains strongly significant for all horizons examined. This evidence is in agreement with the results of recent studies casting doubt on the prevailing view that predictability becomes stronger as the horizon increases (see inter alia Ang and Bekaert, 2007, and Boudoukh et al., 2008).

Concluding, the proposed IVX estimation methodology improves testing in predictive regressions both by extending the range of testable hypotheses and by robustifying inference with respect to misspecification of regressors' persistence. This novel econometric methodology can prove useful for predictability tests in other asset classes too. Successful implementation can shed new light on whether bond yields and exchange rate fluctuations are predictable via publicly available information. Since predictability tests in these asset classes also rely on persistent regressors with uncertain order of integration, this robust methodology can minimize the risk of distorted inference due to incorrect time series modelling.

Appendix: Asymptotic mixed Gaussianity of the IVX estimator

This Appendix provides a summary and discussion of the asymptotic behavior of the normalized and centred IVX estimator $\tilde{A}_{IVX}$ in (17) and $\tilde{A}_{IVX}(K)$ in (33) arising from short-horizon and long-horizon predictive regressions, respectively. The key property of $\tilde{A}_{IVX}$ and $\tilde{A}_{IVX}(K)$ that ensures robustness of the IVX procedure and a chi-squared limit distribution for the IVX-Wald test statistic is asymptotic mixed normality. Theorem A below shows that asymptotic mixed normality applies to all predictors belonging to the classes P(i)-P(iv) of autoregressive processes irrespective of their persistence properties. Theorem B shows that the asymptotic mixed normality property of the IVX estimator extends to long-horizon predictive regression systems. We employ the shorthand notation $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$.

Theorem A. *Consider the model (1)–(3) under Assumption INNOV with instruments* $\tilde{z}_t$ *defined by (4) and (5). Let* B_u *be a r-variate Brownian motion with covariance matrix* Ω_{uu}*,* $J_C(t) = \int_0^t e^{C(t-s)} dB_u(s)$ *be an Ornstein-Uhlenbeck process and let*

$$\underline{B}_u(t) = B_u(t) - \int_0^1 B_u(t)\,dt, \;\; \underline{J}_C(t) = J_C(t) - \int_0^1 J_C(t)\,dt$$

denote the demeaned versions of B_u *and* J_C*. The following limit theory as* $n \to \infty$ *applies for the estimator* $\tilde{A}_{IVX}$ *in (17):*

(i) *when* $\beta < \alpha \wedge 1$, $n^{\frac{1+\beta}{2}} \text{vec}\left(\tilde{A}_{IVX} - A\right) \Rightarrow MN\left(0, \left(\tilde{\Psi}_{uu}^{-1}\right)' C_z V_{C_z} C_z \tilde{\Psi}_{uu}^{-1} \otimes \Sigma_{\varepsilon\varepsilon}\right)$

(ii) *when* $\alpha \in (0, \beta)$, $n^{\frac{1+\alpha}{2}} \text{vec}\left(\tilde{A}_{IVX} - A\right) \Rightarrow N\left(0, V_C^{-1} \otimes \Sigma_{\varepsilon\varepsilon}\right)$

(iii) *when* $\alpha = \beta > 0$, $n^{\frac{1+\alpha}{2}} \text{vec}\left(\tilde{A}_{IVX} - A\right) \Rightarrow N\left(0, \mathbb{V}^{-1} C^{-1} V_C C^{-1} \left(\mathbb{V}'\right)^{-1} \otimes \Sigma_{\varepsilon\varepsilon}\right)$

(iv$_a$) *when* $\alpha = 0$, $\sqrt{n}\text{vec}\left(\tilde{A}_{IVX} - A\right) \Rightarrow N\left(0, \left(E x_{0,1} x_{0,1}'\right)^{-1} \otimes \Sigma_{\varepsilon\varepsilon}\right)$ *under INNOV(i)*

(iv$_b$) *when* $\alpha = 0$, $\sqrt{n}\text{vec}\left(\tilde{A}_{IVX} - A\right) \Rightarrow N\left(0, V_0\right)$ *under INNOV(ii)*

where $x_{0,t} = \sum_{j=0}^{\infty} R^j u_{t-j}$ *with* $R = I_r + C$ *is a stationary version of* x_t *when* $\alpha = 0$*, the covariance matrices* V_C, V_{C_z}, $\mathbb{V}$ *and* V_0 *are given by*

$$\begin{aligned} V_C &= \int_0^\infty e^{rC} \Omega_{uu} e^{rC} dr, \; V_{C_z} = \int_0^\infty e^{rC_z} \Omega_{uu} e^{rC_z} dr, \; \mathbb{V} = \int_0^\infty e^{rC} V_C e^{rC_z} dr, \\ V_0 &= \left(\left[E x_{0,1} x_{0,1}'\right]^{-1} \otimes I_m\right) E\left(x_{0,1} x_{0,1}' \otimes \varepsilon_2 \varepsilon_2'\right) \left(\left[E x_{0,1} x_{0,1}'\right]^{-1} \otimes I_m\right) \end{aligned} \quad (35)$$

and the random covariance matrix $\tilde{\Psi}_{uu}$ *is given by*

$$\tilde{\Psi}_{uu} = \begin{cases} \Omega_{uu} + \int_0^1 \underline{B}_u dB_u' & \text{under } P(i) \\ \Omega_{uu} + \int_0^1 \underline{J}_C dJ_C' & \text{under } P(ii) \\ \Omega_{uu} + V_C C & \text{under } P(iii). \end{cases}$$

The proof of Theorem A can be found in the Online Appendix.

Remarks A.

(a) Theorem A establishes asymptotic mixed normality of the IVX estimator in predictive regression systems the validity of which is invariant to the persistence properties of the generating mechanism of the predictor process x_t. The fact that asymptotic mixed normality extends over the entire range P(i)-P(iv) of autoregression-induced persistence is the key property that ensures robustness of the IVX procedure. The varying rates of convergence and expressions of the (possibly random) limit variance of the IVX estimator along different persistence classes do not affect self-normalized test statistics such as that of the Wald test considered in (19): mixed normality will deliver standard chi-squared asymptotic inference for IVX based self-normalized quadratic forms.

(b) Theorem A shows that the presence of an intercept in the model does not affect the main asymptotic property of IVX estimation, mixed Gaussianity. This, however, is a first-order asymptotic result. In finite samples, the effect of estimating the intercept may become manifest for predictor processes x_t exhibiting high degree of persistence and strong correlation with the innovations ε_t of the predictive equation (1). In this case, represented by part (i) of Theorem A, the sample moment that drives mixed normality can be written as

$$n^{-\frac{1+\beta}{2}} \underline{\mathcal{E}}' \tilde{Z} = n^{-\frac{1+\beta}{2}} \sum_{t=1}^{n} \varepsilon_t \tilde{z}_{t-1}' - n^{\frac{1-\beta}{2}} \bar{\varepsilon}_n \bar{z}_{n-1}'.$$

The first term on the right hand side has a $N\left(0, V_{C_z}^{-1} \otimes \Sigma_{\varepsilon\varepsilon}\right)$ limit distribution which produces the mixed normal limit result for the IVX estimator in part (i) of Theorem A. Using part (i) of Lemma A1 in the Online Appendix, the second term can be analyzed as follows:

$$n^{\frac{1-\beta}{2}}\bar{\varepsilon}_n\bar{z}_{n-1}' = \frac{-C_z^{-1}}{n^{\frac{1-\beta}{2}}n^{\frac{1-\alpha}{2}}}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t\right)\frac{x_n'}{n^{\alpha/2}} = O_p\left(n^{-\frac{1-\beta}{2}}n^{-\frac{1-\alpha}{2}}\right) \tag{36}$$

We conclude that the term in (36) is asymptotically negligible across the whole range P(i)-P(iv) of predictor processes but its finite-sample contribution depends simultaneously on three factors: the degree of regressor persistence α; the correlation between innovations ε_t and u_t; and the choice of β in the instrumentation procedure. The finite-sample impact of the remainder term in (36) is more prominent for highly persistent regressors: persistence of the unit root and local-to-unity type P(i) and P(ii) results to a finite-sample contribution of exact order $O_p\left(n^{-\frac{1-\beta}{2}}\right)$ in (36); the magnitude of this finite-sample contribution declines continuously as the persistence parameter α drives the predictor process towards stationarity and assumes the minimal rate $O_p\left(n^{-1+\beta/2}\right)$ for stationary predictors belonging to the class P(iv). Strong (positive or negative) correlation also exacerbates the finite-sample effect of estimating the intercept in (1): by a simple application of the central limit theorem to (36), it is clear that a unit root predictor $x_n = \sum_{t=1}^{n} u_t$ induces finite-sample bias of the form $-C_z^{-1}n^{-\frac{1-\beta}{2}}\Omega_{\varepsilon u}$, the magnitude of which depends on the long-run covariance $\Omega_{\varepsilon u}$ between the innovations of (1) and (2), defined in (12). All finite-sample effects (irrespective of their source) are simultaneously removed by the finite-sample correction (20) on the self-normalizing component of the IVX-Wald statistic. This correction employs a weighted demeaning of the IVX instruments by a matrix that depends on $\hat{\Omega}_{\varepsilon u}$ in a way that balances the finite-sample contribution of (36) for all persistence and correlation combinations conforming to P(i)-P(iv) and Assumption INNOV and all admissible choices of the IVX tuning parameter β.

Theorem B. *Consider the model (1)–(3) under Assumption INNOV with (9) and Assumption H. The limit distribution as $n \to \infty$ of the normalized and centred long-horizon IVX estimator in (33) is mixed Gaussian of the following form:*

(i) *when* $K/n^{\alpha\wedge\beta} \to 0$, $n^{\frac{1+(\alpha\wedge\beta)}{2}}\mathrm{vec}\left[\tilde{A}_{IVX}(K) - A\right] \Rightarrow MN\left(0, Q_1 \otimes \Sigma_{\varepsilon\varepsilon}\right)$,
$Q_1 = \left(\tilde{\Psi}_{uu}^{-1}\right)' C_z V_{C_z} C_z \tilde{\Psi}_{uu}^{-1}$ *if* $\beta < \alpha$; $Q_1 = V_C^{-1}$ *if* $\alpha < \beta$

(ii) *when* $K/n^{\alpha} \to 0$, $K/n^{\beta} \to \infty$, $\sqrt{nK}\mathrm{vec}\left[\tilde{A}_{IVX}(K) - A\right] \Rightarrow MN(0, Q_2 \otimes \Sigma_{\varepsilon\varepsilon})$,
$Q_2 = \left(\tilde{\Psi}_{uu}^{-1}\right)' \Omega_{uu}^{-1} \tilde{\Psi}_{uu}^{-1}$

(iii) *when* $K/n^{\alpha} \to \infty$, $K/n^{\beta} \to 0$, $\sqrt{n/K}n^{\alpha}\mathrm{vec}\left[\tilde{A}_{IVX}(K) - A\right] \Rightarrow N(0, Q_3 \otimes \Sigma_{\varepsilon\varepsilon})$,
$Q_3 = CV_C^{-1}C^{-1}\Omega_{uu}C^{-1}V_C^{-1}C$ *if* $\alpha > 0$*;* $Q_3 = \left(G_{x_0,\infty}^{-1}\right)' C^{-1}\Omega_{uu}C^{-1}G_{x_0,\infty}^{-1}$ *if* $\alpha = 0$

(iv) *when* $K/n^{\alpha\vee\beta} \to \infty$, $n^{1/2+\alpha-(\alpha\vee\beta)/2}\mathrm{vec}\left[\tilde{A}_{IVX}(K) - A\right] \Rightarrow N(0, Q_4 \otimes \Sigma_{\varepsilon\varepsilon})$,
$Q_4 = 2V_C^{-1}$ *if* $\beta < \alpha$*;* $Q_4 = 2CV_C^{-1}C^{-1}V_{C_z}C^{-1}V_C^{-1}C$ *if* $0 < \alpha < \beta$*; if* $\alpha = 0$ $Q_4 = 2\left(G_{x_0,\infty}^{-1}\right)' C^{-1}V_{C_z}C^{-1}G_{x_0,\infty}^{-1}$.

When $\alpha = 0$ *and* K *is fixed:*

(v_a) $\sqrt{n}\mathrm{vec}\left[\tilde{A}_{IVX}(K) - A\right] \Rightarrow N(0, Q_5 \otimes \Sigma_{\varepsilon\varepsilon})$ *under INNOV(i),*
$Q_5 = \left(G_{x_0,K}^{-1}\right)' \sum_{i,j=0}^{K-1} \Gamma_{x_0}(i-j)\, G_{x_0,K}^{-1}$

(v_b) $\sqrt{n}\mathrm{vec}\left[\tilde{A}_{IVX}(K) - A\right] \Rightarrow N\left(0, \left(G_{x_0,K}^{-1\prime} \otimes I_m\right) W_{0,K} \left(G_{x_0,K}^{-1} \otimes I_m\right)\right)$ *under INNOV(ii),*
$W_{0,K} = \sum_{i,j=0}^{K-1} E\left(x_{0,i}x_{0,j}' \otimes \varepsilon_K \varepsilon_K'\right)$

where V_C, V_{C_z} *and* $\tilde{\Psi}_{uu}$ *are defined in Theorem A,* $\Gamma_{x_0}(j) = E\left(x_{0,t}x_{0,t-j}'\right)$ *is the autocovariance function of the process* $x_{0,t}$ *defined in Theorem A, and* $G_{x_0,K} = \sum_{j=0}^{K-1} \Gamma_{x_0}(j)$.

The proof of Theorem B requires the development of a new limit theory for sample moments arising from long-horizon predictive regressions and joint control of the asymptotic growth rates for n^{α}, n^{β} and K. The details of this asymptotic development are lengthy and highly non-trivial and can be found in Kostakis, Magdalinos and Stamatogiannis (2014).

References

[1] Amihud, Y., Hurvich, C.M., 2004. Predictive regression: A reduced-bias estimation method. Journal of Financial and Quantitative Analysis 39, 813-841.

[2] Amihud, Y., Hurvich, C.M., Wang, Y., 2009. Multiple-predictor regressions: Hypothesis testing. Review of Financial Studies 22, 413-434.

[3] Ang, A., Bekaert, G., 2007. Stock return predictability: Is it there?. Review of Financial Studies 20, 651-707.

[4] van Binsbergen, J., Koijen, R., 2010. Predictive regressions: A present-value approach. Journal of Finance 65, 1439-1471.

[5] Boudoukh, J., Richardson, M.P., Whitelaw, R.F., 2008. The Myth of Long-Horizon Predictability. Review of Financial Studies 21, 1577-1605.

[6] Boussama, F., 2006. Ergodicité des chaînes de Markov à valeurs dans une variété algébrique: application aux modèles GARCH multivariés. Comptes Rendus de l' Academie des Sciences Paris, 343, 275-278.

[7] Campbell, J.Y., 2000. Asset Pricing at the Millenium. Journal of Finance 55, 1515-1567.

[8] Campbell, J.Y, Shiller, R.J., 1988. Stock prices, earnings, and expected dividends. Journal of Finance 43, 661-676.

[9] Campbell, J.Y., Vuolteenaho, T., 2004. Bad beta, good beta. American Economic Review 94, 1249-1275.

[10] Campbell J.Y.,Yogo, M., 2006. Efficient tests of stock return predictability. Journal of Financial Economics 81, 27-60.

[11] Cavanagh, C., Elliott, G., Stock, J.H., 1995. Inference in models with nearly integrated regressors. Econometric Theory 11, 1131-1147.

[12] Cochrane, J., 1999. New facts in finance. Economic Perspectives 23, 36-58.

[13] Cochrane, J., 2008. The dog that did not bark: A defense of return predictability. Review of Financial Studies 21, 1533-1575.

[14] Cochrane, J., 2011. Presidential Address: Discount rates. Journal of Finance 66, 1047-1108.

[15] Elliott, G., 1998. On the robustness of cointegration methods when regressors almost have unit roots. Econometrica 66, 149-158.

[16] Elliott, G., Rothenberg, T.J., Stock, J.H., 1996. Efficient tests for an autoregressive unit root. Econometrica 64, 813-836.

[17] Fama, E.F., 1970. Efficient capital markets: A review of theory and empirical work. Journal of Finance 25, 383-417.

[18] Fama, E.F., 1991. Efficient capital markets: II. Journal of Finance 46, 1575-1617.

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

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

[21] Ferson, W.E., Schadt, R.W., 1996. Measuring fund strategy and performance in changing economic conditions. Journal of Finance 51, 425-461.

[22] Ferson, W.E., Sarkissian, S., Simin, T., 2008. Asset pricing models with conditional betas and alphas: The effects of data snooping and spurious regression. Journal of Financial and Quantitative Analysis 43, 331-354.

[23] Francq, C., Zakoian, JM., 2010. GARCH Models: stucture, statistical inference and financial applications. Wiley.

[24] Giraitis, L., Kokoszka, P., Leipus R., 2000. Stationary ARCH models: dependence structure and the central limit theorem. Econometric Theory, 3-22.

[25] Golez, B., 2014. Expected returns and dividend growth rates implied by derivative markets. Review of Financial Studies, forthcoming.

[26] Hansen, L., Hodrick, R.J., 1980. Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. Journal of Political Economy 88, 829–853.

[27] Hjalmarsson, E., 2011. New methods for inference in long-horizon regressions. Journal of Financial and Quantitative Analysis 46, 815-839.

[28] Hodrick, R.J., 1992. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. Review of Financial Studies 5, 357–386.

[29] Jansson, M., Moreira, M.J., 2006. Optimal inference in regression models with nearly integrated regressors. Econometrica 74, 681-714.

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

[31] Kelly, B., Pruitt, S., 2013. Market expectations in the cross section of present values. Journal of Finance 68, 1721-1756.

[32] Kostakis, A., Magdalinos, T., Stamatogiannis, 2014. IVX limit theory for long-horizon predictive regressions. Unpublished manuscript, University of Southampton.

[33] Kothari, S., Shanken, J., 1997. Book-to-market, dividend yield, and expected market returns: a time-series analysis. Journal of Financial Economics 44, 169-203.

[34] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationary against the alternative of a unit root. Journal of Econometrics 54, 159-178.

[35] Lamont, O., 1998. Earnings and expected returns. Journal of Finance 53, 1563-1587.

[36] Lanne, M., 2002. Testing the predictability of stock returns. Review of Economics and Statistics 84, 407-415.

[37] Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. Journal of Finance 56, 815-849.

[38] Lettau, M., Ludvigson, S., 2005. Expected returns and expected dividend growth. Journal of Financial Economics 76, 583-626.

[39] Lettau, M., van Nieuwerburgh, S., 2008. Reconciling the return predictability evidence. Review of Financial Studies 21, 1607-1652.

[40] Magdalinos, T., Phillips, P.C.B., 2009. Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. Econometric Theory 25, 482-526.

[41] Newey, W., West, K., 1987. A simple, positive definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

[42] Phillips, P.C.B., 1987. Time series regression with a unit root. Econometrica 55, 277-302.

[43] Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regression with I(1) processes. Review of Economic Studies 57, 99-125.

[44] Phillips, P.C.B., Magdalinos, T., 2007. Limit theory for Moderate deviations from a unit root. Journal of Econometrics 136, 115-130.

[45] Phillips, P.C.B., Magdalinos T., 2009. Econometric inference in the vicinity of unity. CoFie Working Paper No. 7, Singapore Management University.

[46] Phillips, P.C.B., Solo, V., 1992. Asymptotics for Linear Processes. Annals of Statistics 20, 971-1001.

[47] Saikkonen, P., 1991. Asymptotically efficient estimation of cointegration regressions. Econometric Theory 7, 1-21.

[48] Stambaugh, R.F., 1999. Predictive regressions. Journal of Financial Economics 54, 375-421.

[49] Stock, J.H., Watson, M.W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. Econometrica 61, 783-820.

[50] Torous, W., Valkanov, R., Yan, S., 2004. On predicting stock returns with nearly integrated explanatory variables. Journal of Business 77, 937-966.

[51] Valkanov, R., 2003. Long-horizon regressions: Theoretical results and applications. Journal of Financial Economics 68, 201-232.

[52] Wei, M., Wright, J.H., 2013. Reverse regressions and long-horizon forecasting. Journal of Applied Econometrics 28, 353-371.

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

[54] White, H., 1980. A Heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817-839.

Table 1

Finite−sample sizes when there is no autocorrelation ($\phi = 0$) in the residuals of the autoregression

This table presents finite−sample sizes, testing the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) when there is no autocorrelation in the residuals of the autoregressive equation, i.e., $\phi = 0$ in (24). $W_{0.05}$ corresponds to the rejection rate for the Wald statistic, defined in (19), with 5% nominal size, $Q_{0.05}$ corresponds to the rejection rate resulting from the 95% confidence interval for the Campbell and Yogo (2006) *Q*-test and $JM_{0.05}$ corresponds to the rejection rate for the $\pi^*_{0.05}$ statistic of Jansson and Moreira (2006). Results are reported for different degrees of correlation between the residuals of regressions (22) and (23), $\delta = -0.95$, −0.5, 0, 0.5 and 0.95, different sample sizes *n*= 100, 250, 500 and 1,000 and for different local-to-unity parameters *C*= 0, −5, −10, −15, −20 and −50, which in each sample size case correspond to different autoregressive roots (R_n) reported in the third column. The reported results are based on the Monte Carlo simulation described in Section 2.1 and the average rejection rates are calculated over 10,000 repetitions.

			$\delta = -0.95$			$\delta = -0.50$			$\delta = 0$			$\delta = 0.50$			$\delta = 0.95$		
n	*C*	R_n	$W_{0.05}$	$Q_{0.05}$	$JM_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$JM_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$JM_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$JM_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$JM_{0.05}$
100	0	1.000	0.067	0.055	0.048	0.064	0.044	0.062	0.051	0.050	0.436	0.063	0.042	0.060	0.063	0.054	0.058
	−5	0.950	0.072	0.061	0.046	0.060	0.039	0.046	0.055	0.050	0.192	0.057	0.037	0.052	0.070	0.062	0.044
	−10	0.900	0.066	0.068	0.030	0.060	0.039	0.032	0.059	0.052	0.170	0.056	0.039	0.040	0.065	0.064	0.028
	−20	0.800	0.063	0.088	0.066	0.056	0.044	0.068	0.051	0.045	0.144	0.057	0.042	0.068	0.062	0.085	0.070
	−50	0.500	0.058	0.257	0.150	0.050	0.095	0.058	0.058	0.054	0.148	0.054	0.094	0.048	0.055	0.257	0.148
250	0	1.000	0.060	0.051	0.062	0.053	0.036	0.054	0.050	0.050	0.510	0.057	0.038	0.042	0.057	0.046	0.052
	−5	0.980	0.062	0.047	0.036	0.056	0.034	0.048	0.050	0.050	0.208	0.052	0.031	0.038	0.062	0.046	0.028
	−10	0.960	0.059	0.050	0.042	0.055	0.032	0.052	0.051	0.048	0.158	0.048	0.030	0.036	0.061	0.053	0.042
	−20	0.920	0.057	0.062	0.040	0.050	0.032	0.036	0.052	0.049	0.128	0.054	0.033	0.038	0.059	0.059	0.034
	−50	0.800	0.054	0.169	0.318	0.050	0.050	0.038	0.055	0.052	0.116	0.053	0.054	0.040	0.055	0.166	0.342
500	0	1.000	0.052	0.039	0.042	0.053	0.038	0.046	0.049	0.048	0.582	0.051	0.036	0.072	0.059	0.043	0.048
	−5	0.990	0.062	0.049	0.036	0.051	0.030	0.038	0.051	0.048	0.258	0.052	0.032	0.040	0.064	0.050	0.040
	−10	0.980	0.057	0.044	0.036	0.055	0.031	0.036	0.049	0.049	0.200	0.054	0.033	0.040	0.060	0.047	0.032
	−20	0.960	0.055	0.049	0.050	0.054	0.029	0.042	0.051	0.051	0.178	0.049	0.028	0.048	0.056	0.049	0.054
	−50	0.900	0.052	0.113	0.524	0.052	0.037	0.054	0.048	0.045	0.176	0.051	0.037	0.054	0.054	0.114	0.488
1000	0	1.000	0.055	0.042	0.038	0.047	0.034	0.038	0.051	0.049	0.646	0.052	0.035	0.032	0.056	0.042	0.046
	−5	0.995	0.059	0.047	0.040	0.051	0.030	0.046	0.052	0.051	0.334	0.055	0.031	0.034	0.061	0.048	0.042
	−10	0.990	0.059	0.046	0.038	0.052	0.030	0.050	0.051	0.048	0.270	0.054	0.032	0.050	0.055	0.046	0.046
	−20	0.980	0.058	0.047	0.042	0.057	0.031	0.034	0.049	0.047	0.222	0.053	0.029	0.040	0.060	0.048	0.036
	−50	0.950	0.052	0.074	0.606	0.050	0.032	0.028	0.049	0.048	0.194	0.049	0.029	0.030	0.056	0.069	0.600

Table 2

Finite−sample sizes with autocorrelation coefficient $\phi = 0.5$ in the residuals of the autoregression

This table presents finite−sample sizes, testing the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) when the autocorrelation coefficient in the residuals of the autoregression (23) is $\phi = 0.5$. $W_{0.05}$ corresponds to the rejection rate for the Wald statistic, defined in (19), with 5% nominal size, $Q_{0.05}$ corresponds to the rejection rate resulting from the 95% confidence interval for the Campbell and Yogo (2006) *Q*-test and $JM_{0.05}$ corresponds to the rejection rate for the $\pi^*_{0.05}$ statistic of Jansson and Moreira (2006). Results are reported for different degrees of correlation between the residuals of regressions (22) and (23), $\delta = -0.95$, −0.5, 0, 0.5 and 0.95, different sample sizes *n*= 100, 250, 500 and 1,000 and for different local-to-unity parameters *C*= 0, −5, −10, −15, −20 and −50, which in each sample size case correspond to different autoregressive roots (R_n) reported in the third column. The reported results are based on the Monte Carlo simulation described in Section 2.1 and the average rejection rates are calculated over 10,000 repetitions.

			$\delta = -0.95$			$\delta = -0.50$			$\delta = 0$			$\delta = 0.50$			$\delta = 0.95$		
n	*C*	R_n	$W_{0.05}$	$Q_{0.05}$	$JM_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$JM_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$JM_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$JM_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$JM_{0.05}$
100	0	1.000	0.072	0.054	0.110	0.066	0.044	0.110	0.050	0.051	0.394	0.061	0.039	0.118	0.073	0.054	0.108
	−5	0.950	0.072	0.053	0.148	0.063	0.040	0.056	0.053	0.049	0.162	0.062	0.037	0.050	0.073	0.053	0.136
	−10	0.900	0.068	0.047	0.156	0.060	0.036	0.046	0.054	0.050	0.124	0.061	0.034	0.038	0.071	0.052	0.138
	−20	0.800	0.063	0.059	0.140	0.055	0.032	0.038	0.056	0.051	0.094	0.056	0.033	0.034	0.061	0.056	0.138
	−50	0.500	0.053	0.150	0.134	0.051	0.053	0.042	0.055	0.052	0.100	0.056	0.055	0.046	0.055	0.155	0.112
250	0	1.000	0.064	0.044	0.122	0.055	0.033	0.088	0.051	0.052	0.420	0.054	0.033	0.070	0.060	0.044	0.114
	−5	0.980	0.065	0.046	0.124	0.059	0.033	0.054	0.051	0.048	0.158	0.057	0.034	0.052	0.067	0.045	0.134
	−10	0.960	0.066	0.046	0.118	0.057	0.035	0.044	0.055	0.050	0.108	0.058	0.032	0.038	0.062	0.043	0.116
	−20	0.920	0.054	0.046	0.112	0.056	0.033	0.036	0.049	0.047	0.078	0.058	0.034	0.030	0.056	0.047	0.122
	−50	0.800	0.054	0.150	0.094	0.051	0.044	0.040	0.051	0.048	0.102	0.054	0.046	0.048	0.057	0.144	0.112
500	0	1.000	0.055	0.043	0.070	0.053	0.036	0.052	0.047	0.049	0.410	0.050	0.034	0.072	0.056	0.043	0.088
	−5	0.990	0.064	0.044	0.104	0.056	0.033	0.052	0.052	0.048	0.202	0.056	0.033	0.052	0.062	0.049	0.108
	−10	0.980	0.061	0.044	0.082	0.053	0.032	0.026	0.047	0.044	0.152	0.053	0.030	0.036	0.061	0.044	0.074
	−20	0.960	0.055	0.043	0.114	0.050	0.029	0.040	0.050	0.046	0.136	0.052	0.033	0.042	0.058	0.045	0.102
	−50	0.900	0.051	0.097	0.112	0.049	0.034	0.060	0.056	0.053	0.136	0.050	0.033	0.058	0.057	0.098	0.120
1000	0	1.000	0.054	0.039	0.066	0.056	0.034	0.044	0.052	0.053	0.468	0.054	0.033	0.044	0.061	0.044	0.096
	−5	0.995	0.065	0.049	0.088	0.057	0.035	0.060	0.053	0.053	0.216	0.054	0.030	0.060	0.063	0.046	0.112
	−10	0.990	0.060	0.047	0.096	0.055	0.031	0.062	0.047	0.045	0.146	0.052	0.032	0.054	0.060	0.046	0.106
	−20	0.980	0.061	0.045	0.100	0.053	0.030	0.040	0.051	0.047	0.124	0.051	0.028	0.042	0.064	0.050	0.104
	−50	0.950	0.052	0.064	0.124	0.052	0.027	0.036	0.053	0.051	0.116	0.053	0.028	0.034	0.053	0.064	0.110

Table 3
Finite−sample sizes for multivariate predictive systems

This table presents finite−sample sizes for four Wald tests, with nominal size 5%, based on the multivariate predictive system in (26) with three regressors exhibiting different degrees of persistence (unit root, local-to-unity and stationary), as described in the Monte Carlo simulation in Section 2.4. W_{joint} reports the rejection rate for the joint Wald test, defined in (19), under the null hypothesis $H_0 : A = 0_{1x3}$, i.e., that all three coefficients in vector *A* are equal to zero. W_{UR} reports the corresponding rejection rate for the individual significance of the unit root regressor coefficient, i.e., under the null hypothesis $H_0 : A_1 = 0$. W_{LTU} reports the rejection rate for the individual significance of the local-to-unity regressor coefficient, i.e., under the null hypothesis $H_0 : A_2 = 0$, while $W_{Stationary}$ reports the rejection rate for the individual significance of the stationary regressor coefficient, i.e., under the null hypothesis $H_0 : A_3 = 0$. Results are reported for (i) two sets of correlations (δ's) between the residuals of regressions (26) and (27), as estimated using S&P 500 value-weighted log excess return (regressand), earnings-price ratio (UR), T-bill rate (LTU) and inflation rate (Stationary) with monthly (Correlation Set 1) and quarterly (Correlation Set 2) data for the period 1927−2012, (ii) four sets of autocorrelation coefficients in the residuals of autoregressions in (27): ϕ=0, 0.25, 0.5 and the corresponding sample estimates for each of the three regressors mentioned above and (iii) different sample sizes: n=250, 500 and 1,000. The average rejection rates are calculated over 10,000 repetitions.

Correlation Set 1	*n*	W_{joint}	W_{UR}	W_{LTU}	$W_{Stationary}$
	250	0.052	0.078	0.065	0.057
$\phi_1 = \phi_2 = \phi_3 = 0$	500	0.051	0.076	0.060	0.057
	1000	0.047	0.077	0.065	0.054
	250	0.070	0.076	0.065	0.055
$\phi_1 = \phi_2 = \phi_3 = 0.25$	500	0.064	0.080	0.062	0.053
	1000	0.063	0.075	0.067	0.049
	250	0.058	0.082	0.067	0.053
$\phi_1 = \phi_2 = \phi_3 = 0.5$	500	0.053	0.079	0.069	0.053
	1000	0.049	0.080	0.059	0.052
$\phi_1 = 0.28$	250	0.070	0.084	0.065	0.055
$\phi_2 = 0.32$	500	0.064	0.080	0.069	0.052
$\phi_3 = -0.14$	1000	0.067	0.079	0.062	0.053
Correlation Set 2	***n***	W_{joint}	W_1	W_2	W_3
	250	0.058	0.054	0.058	0.056
$\phi_1 = \phi_2 = \phi_3 = 0$	500	0.048	0.053	0.059	0.050
	1000	0.051	0.054	0.054	0.054
	250	0.057	0.054	0.053	0.058
$\phi_1 = \phi_2 = \phi_3 = 0.25$	500	0.052	0.055	0.050	0.055
	1000	0.056	0.050	0.053	0.051
	250	0.054	0.058	0.057	0.054
$\phi_1 = \phi_2 = \phi_3 = 0.5$	500	0.049	0.058	0.059	0.047
	1000	0.052	0.053	0.054	0.052
$\phi_1 = 0.22$	250	0.058	0.053	0.060	0.052
$\phi_2 = -0.1$	500	0.053	0.053	0.052	0.051
$\phi_3 = -0.08$	1000	0.052	0.052	0.049	0.053

Table 4
Unit root tests for predictive regressors−Monthly data
This table presents the results of unit root tests for the following list of financial and economic variables defined in Section 3: Dividend payout ratio (d/e), long-term yield (lty), dividend yield (d/y), dividend price ratio (d/p), T-bill rate (tbl), earnings price ratio (e/p), book-to-market value ratio (b/m), default yield spread (dfy), net equity expansion (ntis), term spread (tms) and inflation rate (inf). $\hat{R}_n$ corresponds to the least squares point estimate of the AR(1): $x_t = R_n x_{t-1} + u_t$. ADF stands for the augmented Dickey-Fuller test statistic, DF-GLS refers to the Elliot et al. (1996) Dickey-Fuller-GLS test statistic, PP stands for the Phillips-Perron test statistic and KPSS refers to the Kwiatkowski et al. (1992) test statistic. The Bayesian Information Criterion has been been used to select the optimal lag length for ADF and DF-GLS test statistics. The sample period is January 1927−December 2012. *, ** and *** imply rejection of the null hypothesis of a unit root (for ADF, DF-GLS and PP) or stationarity (for KPSS) at 10%, 5% and 1% level respectively.

	$\hat{R}_n$	ADF	DF-GLS	PP	KPSS
Dividend payout ratio	0.999	−5.758***	−5.712***	−4.184***	1.701***
Long-term yield	0.999	−1.286	−1.181	−1.314	1.853***
Dividend yield	1.000	−2.179	−1.448	−2.087	2.502***
Dividend-price ratio	1.000	−2.180	−1.468	−2.149	2.505***
T-bill rate	0.997	−2.238	−2.237**	−2.131	1.313***
Earnings-price ratio	1.000	−3.870***	−3.014***	−3.656***	1.026***
Book-to-market value ratio	0.997	−3.108**	−2.754***	−2.989**	1.384***
Default yield spread	0.993	−3.430**	−3.364***	−3.779***	0.546**
Net equity expansion	0.981	−4.371***	−1.247	−4.592***	1.008***
Term spread	0.985	−5.112***	−3.727***	−4.697***	0.535**
Inflation rate	0.633	−9.161***	−5.257***	−20.531***	0.617**

Table 5

Unit root tests for predictive regressors−Quarterly data

This table presents the results of unit root tests for the following list of financial and economic variables defined in Section 3: Dividend payout ratio (d/e), long-term yield (lty), dividend yield (d/y), dividend price ratio (d/p), T-bill rate (tbl), earnings price ratio (e/p), book-to-market value ratio (b/m), default yield spread (dfy), net equity expansion (ntis), term spread (tms), inflation rate (inf) and consumption-wealth ratio (cay). $\hat{R}_n$ corresponds to the least squares point estimate of the AR(1): $x_t = R_n x_{t-1} + u_t$. ADF stands for the augmented Dickey-Fuller test statistic, DF-GLS refers to the Elliot et al. (1996) Dickey-Fuller-GLS test statistic, PP stands for the Phillips-Perron test statistic and KPSS refers to the Kwiatkowski et al. (1992) test statistic. The Bayesian Information Criterion has been used to select the optimal lag length for ADF and DF-GLS test statistics. The sample period is 1927Q1−2012Q4, with the exception of cay, which becomes available at quarterly frequency after 1952. *, ** and *** imply rejection of the null hypothesis of a unit root (for ADF, DF-GLS and PP) or stationarity (for KPSS) at 10%, 5% and 1% level respectively.

	$\hat{R}_n$	ADF	DF-GLS	PP	KPSS
Dividend payout ratio	0.985	−4.019***	−3.995***	−3.938***	1.288***
Long-term yield	0.997	−1.428	−1.318	−1.213	1.023***
Dividend yield	1.000	−2.159	−1.560	−2.096	1.439***
Dividend-price ratio	1.000	−2.224	−1.619*	−2.284	1.453***
T-bill rate	0.983	−2.141	−2.145**	−2.333	0.765***
Earnings-price ratio	0.999	−4.274***	−2.462**	−3.424**	0.665**
Book-to-market value ratio	0.989	−3.500***	−3.114***	−3.262**	0.800***
Default yield spread	0.971	−3.241**	−3.186***	−4.055***	0.357*
Net equity expansion	0.939	−4.182***	−1.057	−4.654***	0.752***
Term spread	0.944	−4.536***	−2.923***	−5.333***	0.418*
Inflation rate	0.627	−4.364***	−4.366***	−12.360***	0.425*
Consumption-wealth ratio	0.951	−2.408	−2.201**	−2.431	0.232

Table 6
Univariate predictive regressions−Monthly data

This table presents the results of univariate predictive regression models, as in equation (1), during the sample periods January 1927−December 2012 (Panel A) and January 1952−December 2012 (Panel B). The dependent variable is the monthly S&P 500 value-weighted log excess return and the lagged persistent regressor is each of the following variables defined in Section 3: Dividend payout ratio (d/e), long-term yield (lty), dividend yield (d/y), dividend price ratio (d/p), T-bill rate (tbl), earnings price ratio (e/p), book-to-market value ratio (b/m), default yield spread (dfy), net equity expansion (ntis), term spread (tms) and inflation rate (inf). $\tilde{A}_{OLS}$ stands for the least squares slope coefficient estimated via regression model (1), while t_{OLS} is the corresponding t-statistic under the null hypothesis that A is equal to zero (i.e., no predictability). $\tilde{A}_{IVX}$, defined in (17), stands for the slope coefficient for the predictive regression (16) estimated via the proposed instrumental variable (IVX) approach, while IVX-Wald refers to the Wald statistic, defined in equation (19), under the null hypothesis that the slope coefficient A is equal to zero. δ denotes the correlation coefficient between the residuals of regression models (1) and (2). *, ** and *** imply rejection of the null hypothesis at 10%, 5% and 1% level respectively. CY 90% CI stands for the 90% Bonferroni confidence interval for the bias-corrected scaled least squares slope coefficient of the predictive regression using the *Q*-test of Campbell and Yogo (2006). Bold fonts indicate rejection of the null hypothesis of no predictability at the 10% level. JM reports the p-value for the $\pi^{*}_{0.05}$ statistic of Jansson and Moreira (2006) under the null hypothesis of no predictability.

Regressors	$\tilde{A}_{OLS}$	t_{OLS}	$\tilde{A}_{IVX}$	IVX−Wald	δ	CY 90% CI		JM
Panel A: January 1927−December 2012								
Dividend payout ratio	−0.0024	−0.46	−0.0033	0.393	−0.067	−0.006	0.003	0.19
Long-term yield	−0.0622	−1.01	−0.0665	1.064	−0.108	−0.007	0.002	0.38
Dividend yield	0.0075	1.97**	0.0081	3.129*	−0.079	**0.001**	**0.014**	0.06*
Dividend-price ratio	0.0062	1.63	0.0065	2.031	−0.975	−0.004	0.008	0.32
T-bill rate	−0.0784	−1.40	−0.0761	1.770	−0.062	−0.011	0.001	0.03**
Earnings-price ratio	0.0087	2.13**	0.0088	4.402**	−0.759	−0.003	0.015	0.34
Book-to-market value ratio	0.0148	2.28**	0.0134	4.101**	−0.823	**0.001**	**0.021**	0.12
Default yield spread	0.1100	0.45	0.0591	0.058	−0.274	−0.009	0.015	0.03**
Net equity expansion	−0.1355	−1.93*	−0.1720	4.150**	−0.031	**−0.026**	**−0.003**	0.01***
Term spread	0.1482	1.13	0.1399	1.095	−0.005	−0.004	0.024	0.15
Inflation rate	−0.3500	−1.07	−0.3555	1.148	0.023	−0.064	0.021	0.35
Panel B: January 1952−December 2012								
Dividend payout ratio	0.0049	0.93	0.0044	0.672	−0.091	−0.003	0.009	0.31
Long-term yield	−0.0725	−1.23	−0.0777	1.396	−0.148	−0.012	0.002	0.16
Dividend yield	0.0075	1.95*	0.0081	1.425	−0.058	**0.001**	**0.014**	0.04**
Dividend-price ratio	0.0069	1.79*	0.0072	1.142	−0.986	−0.006	0.005	0.43
T-bill rate	−0.1057	−2.01**	−0.1054	3.537*	−0.126	**−0.018**	**−0.002**	0.27
Earnings-price ratio	0.0038	1.04	0.0029	0.588	−0.610	−0.011	0.006	0.46
Book-to-market value ratio	0.0043	0.68	0.0029	0.174	−0.747	−0.007	0.008	0.27
Default yield spread	0.2275	0.65	0.2306	0.389	−0.056	−0.009	0.019	0.46
Net equity expansion	−0.0259	−0.30	−0.0417	0.220	−0.063	−0.016	0.010	0.28
Term spread	0.2071	1.88*	0.2176	3.808*	0.034	**0.002**	**0.038**	0.03**
Inflation rate	−1.0501	−2.31**	−1.1057	5.922**	−0.069	**−0.130**	**−0.031**	0.15

Table 7
Univariate predictive regressions−Quarterly data
This table presents the results of univariate predictive regression models, as in equation (1), during the sample period 1927Q1−2012Q4 (Panel A) and 1952Q1−2012Q4 (Panel B). The dependent variable is the quarterly S&P 500 value-weighted log excess return and the lagged persistent regressor is each of the following variables defined in Section 3: Dividend payout ratio (d/e), long-term yield (lty), dividend yield (d/y), dividend price ratio (d/p), T-bill rate (tbl), earnings price ratio (e/p), book-to-market value ratio (b/m), default yield spread (dfy), net equity expansion (ntis), term spread (tms), inflation rate (inf) and consumption-wealth ratio (cay). $\tilde{A}_{OLS}$ stands for the least squares slope coefficient estimated via regression model (1), while t_{OLS} is the corresponding t-statistic under the null hypothesis that A is equal to zero (i.e., no predictability). $\tilde{A}_{IVX}$, defined in (17), stands for the slope coefficient for the predictive regression (16) estimated via the proposed instrumental variable (IVX) approach, while IVX-Wald refers to the Wald statistic, defined in equation (19), under the null hypothesis that the slope coefficient A is equal to zero. δ denotes the correlation coefficient between the residuals of regression models (1) and (2). *, ** and *** imply rejection of the null hypothesis at 10%, 5% and 1% level respectively. CY 90% CI stands for the 90% Bonferroni confidence interval for the bias-corrected scaled least squares slope coefficient of the predictive regression using the *Q*-test of Campbell and Yogo (2006). Bold fonts indicate rejection of the null hypothesis of no predictability at the 10% level. JM reports the p-value for the $\pi^{*}_{0.05}$ statistic of Jansson and Moreira (2006) under the null hypothesis of no predictability.

Regressors	$\tilde{A}_{OLS}$	t_{OLS}	$\tilde{A}_{IVX}$	IVX−Wald	δ	CY 90% CI		JM
Panel A: 1927Q1−2012Q4								
Dividend payout ratio	−0.0031	−0.18	−0.0053	0.095	−0.138	−0.037	0.020	0.22
Long-term yield	−0.1621	−0.78	−0.1705	0.629	−0.071	−0.022	0.008	0.34
Dividend yield	0.0216	1.69*	0.0232	2.638	0.045	**0.001**	**0.044**	0.03**
Dividend-price ratio	0.0230	1.83*	0.0249	2.952*	−0.943	−0.010	0.033	0.35
T-bill rate	−0.2110	−1.13	−0.2032	1.129	−0.029	−0.039	0.008	0.07*
Earnings-price ratio	0.0284	2.10**	0.0289	4.439**	−0.556	−0.002	0.072	0.31
Book-to-market value ratio	0.0610	2.82***	0.0565	6.553**	−0.832	−0.001	0.062	0.10*
Default yield spread	0.6472	0.80	0.5041	0.390	−0.515	−0.026	0.064	0.01**
Net equity expansion	−0.6054	−2.60***	−0.7683	6.596**	0.137	**−0.090**	**−0.022**	0.04**
Term spread	0.4245	0.97	0.4007	0.796	−0.005	−0.016	0.076	0.17
Inflation rate	−0.1980	−0.45	−0.1954	0.198	0.033	−0.084	0.061	0.43
Panel B: 1952Q1−2012Q4								
Dividend payout ratio	0.0189	1.13	0.0177	1.097	−0.190	−0.024	0.057	0.50
Long-term yield	−0.1792	−0.93	−0.1881	0.782	−0.095	−0.035	0.009	0.14
Dividend yield	0.0272	2.17**	0.0307	2.235	−0.095	**0.004**	**0.046**	0.03**
Dividend-price ratio	0.0237	1.88*	0.0257	1.525	−0.967	−0.016	0.019	0.44
T-bill rate	−0.2835	−1.65*	−0.2806	2.362	−0.073	−0.067	0.001	0.24
Earnings-price ratio	0.0112	0.95	0.0088	0.518	−0.334	−0.028	0.044	0.49
Book-to-market value ratio	0.0200	0.97	0.0171	0.546	−0.793	−0.020	0.028	0.31
Default yield spread	0.6762	0.60	0.6910	0.329	−0.174	−0.041	0.065	0.49
Net equity expansion	−0.0319	−0.11	−0.0718	0.060	−0.034	−0.043	0.344	0.43
Term spread	0.6047	1.68*	0.6349	3.057*	0.040	**0.001**	**0.119**	0.05**
Inflation rate	−0.7879	−1.38	−0.8793	2.356	−0.128	**−0.193**	**−0.026**	0.15
Consumption-wealth ratio	0.8480	3.38***	0.8746	11.351***	−0.429	**0.032**	**0.110**	0.02**

Table 8

Predictive regressions with multiple regressors−Monthly data

This table presents the results of predictive regression models with multiple regressors, as in equation (1), during the sample periods January 1927−December 2012 (Panel A) and January 1952−December 2012 (Panel B). In each case, the dependent variable is the monthly S&P 500 value-weighted log excess return and the lagged regressors are combinations of the following variables defined in Section 3: Dividend price ratio (d/p), earnings price ratio (e/p), book-to-market value ratio (b/m), dividend payout ratio (d/e), T-bill rate (tbl), default yield spread (dfy) and term spread (tms). $\tilde{A}_{IVX}$, defined in (17), is the vector containing the slope coefficients with respect to each of the employed variables for the predictive regression (16), estimated via the instrumental variable (IVX) approach. The significance of each individual coefficient is evaluated using the Wald statistic, defined in equation (19), under the null hypothesis that the corresponding coefficient is equal to zero. Joint Wald refers to the same Wald statistic, under the null hypothesis that all coefficients *A* are jointly equal to zero. *, ** and *** imply rejection of the null hypothesis at 10%, 5% and 1% level respectively.

d/p	e/p	b/m	d/e	tbl	dfy	tms	Joint Wald	Related study/ Model
Panel A: January 1927−December 2012								
0.0061	...	...	...	−0.0807	...	...	3.644	Ang and Bekaert (2007)
0.0077	...	...	...	−0.0647	−0.1871	0.0996	4.742	Ferson and Schadt (1996)
−0.0010	...	0.0150	...	...	...	...	4.117	Kothari and Shanken (1997)
0.0091*	...	...	−0.0082	...	...	...	3.655	Lamont (1998)
...	0.0082	0.0053	...	...	...	0.1992	7.321*	Campbell and Vuolteenaho (2004)
...	0.0112**	...	...	−0.1275**	...	...	8.748**	General-to-specific approach
Panel B: January 1952−December 2012								
d/p	e/p	b/m	d/e	tbl	dfy	tms	Joint Wald	Related study/ Model
0.0150	...	...	...	−0.2314**	...	...	4.132	Ang and Bekaert (2007)
0.0130	...	...	...	−0.2044	0.2252	0.0607	7.653	Ferson and Schadt (1996)
0.0237	...	−0.0290	...	...	...	...	2.085	Kothari and Shanken (1997)
0.0067	...	...	0.0025	...	...	...	1.326	Lamont (1998)
...	0.0060	−0.0014	...	...	...	0.2633**	5.420	Campbell and Vuolteenaho (2004)
...	0.0108**	...	...	−0.2113***	...	...	8.160**	General-to-specific approach

Table 9

Predictive regressions with multiple regressors−Quarterly data

This table presents the results of predictive regression models with multiple regressors, as in equation (1), during the sample periods 1927Q1−2012Q4 (Panel A) and 1952Q1−2012Q4 (Panel B). In each case, the dependent variable is the quarterly S&P 500 value-weighted log excess return and the lagged persistent regressors are combinations of the following variables defined in Section 3: Dividend price ratio (d/p), earnings price ratio (e/p), book-to-market value ratio (b/m), dividend payout ratio (d/e), T-bill rate (tbl), default yield spread (dfy), term spread (tms), consumption-wealth ratio (cay) and net equity expansion (ntis). $\tilde{A}_{IVX}$, defined in (17), is the vector containing the slope coefficients with respect to each of the employed variables for the predictive regression (16), estimated via the instrumental variable (IVX) approach. Joint Wald refers to the Wald statistic, defined in equation (19), under the null hypothesis that all coefficients *A* are jointly equal to zero. *, ** and *** imply rejection of the null hypothesis at 10%, 5% and 1% level respectively.

Panel A: 1927Q1−2012Q4									
d/p	e/p	b/m	d/e	tbl	dfy	tms	ntis	Joint Wald	Related study/ Model
0.0240*	...	...	...	−0.2190	...	...	...	3.971	Ang and Bekaert (2007)
0.0267	...	...	...	−0.1731	−0.2871	0.2476	...	4.557	Ferson and Schadt (1996)
−0.0137	...	0.0770	...	...	...	...	...	6.576**	Kothari and Shanken (1997)
0.0321**	...	...	−0.0222	...	...	...	...	4.023	Lamont (1998)
...	0.0160	0.0413	...	...	...	0.5046	...	8.391**	Campbell and Vuolteenaho (2004)
...	0.0361**	...	...	−0.3755*	...	...	−0.6152*	13.469***	General-to-specific approach

Panel B: 1952Q1−2012Q4									
d/p	e/p	b/m	d/e	tbl	dfy	tms	cay	Joint Wald	Related study/ Model
0.0483	...	...	...	−0.6828*	...	...	...	3.745	Ang and Bekaert (2007)
0.0434	...	...	...	−0.5884	0.5073	0.2380	...	6.880	Ferson and Schadt (1996)
0.0706	...	−0.0783	...	...	...	...	...	1.883	Kothari and Shanken (1997)
0.0235	...	...	0.0114	...	...	...	...	1.954	Lamont (1998)
...	0.0089	0.0161	...	...	...	0.7553**	...	4.574	Campbell and Vuolteenaho (2004)
0.0230	...	...	−0.0006	...	...	...	0.7333**	13.199***	Lettau and Ludvigson (2001)
...	0.0390**	...	...	−0.7339***	2.4016**	...	0.9749***	23.985***	General-to-specific approach

Table 10
Finite−sample sizes for long−horizon Wald test

This table presents finite−sample sizes, derived from K−horizon univariate predictive regressions, as in equation (30), under the null hypothesis $H_0 : A = 0$ in the DGP (22). $W_{0.05}$ corresponds to the rejection rate for the long-horizon Wald statistic, defined in (34), with 5% nominal size. Results are reported for different degrees of correlation between the residuals of regressions (22) and (23), $\delta = -0.95$, −0.5 and 0, different sample sizes n= 100, 500 and 1,000, different horizons K that are empirically relevant to the corresponding sample size n and different local-to-unity parameters C= 0, −5, −10, −20 and −50. The reported results are based on the Monte Carlo simulation described in Section 5.2 and the average rejection rates are calculated over 10,000 repetitions.

n=100					*n*=500					*n*=1000				
		δ=−0.95	*δ=−0.5*	*δ=0*			*δ=−0.95*	*δ=−0.5*	*δ=0*			*δ=−0.95*	*δ=−0.5*	*δ=0*
C	*K*	$W_{0.05}$	$W_{0.05}$	$W_{0.05}$	*C*	*K*	$W_{0.05}$	$W_{0.05}$	$W_{0.05}$	*C*	*K*	$W_{0.05}$	$W_{0.05}$	$W_{0.05}$
0	2	0.067	0.060	0.051	0	4	0.060	0.054	0.050	0	4	0.056	0.055	0.051
	3	0.062	0.062	0.050		8	0.055	0.050	0.051		12	0.053	0.051	0.049
	4	0.059	0.057	0.048		12	0.053	0.050	0.045		36	0.048	0.044	0.048
	5	0.057	0.055	0.047		20	0.050	0.049	0.047		60	0.044	0.042	0.049
−5	2	0.067	0.060	0.053	−5	4	0.060	0.059	0.050	−5	4	0.059	0.056	0.050
	3	0.064	0.060	0.052		8	0.063	0.053	0.053		12	0.060	0.052	0.047
	4	0.062	0.050	0.048		12	0.060	0.052	0.053		36	0.052	0.053	0.045
	5	0.059	0.050	0.048		20	0.057	0.049	0.044		60	0.047	0.043	0.047
−10	2	0.061	0.062	0.050	−10	4	0.059	0.052	0.050	−10	4	0.061	0.049	0.047
	3	0.066	0.057	0.049		8	0.056	0.056	0.050		12	0.054	0.055	0.052
	4	0.059	0.051	0.054		12	0.058	0.054	0.049		36	0.053	0.052	0.048
	5	0.058	0.052	0.047		20	0.053	0.048	0.049		60	0.049	0.044	0.047
−20	2	0.058	0.057	0.055	−20	4	0.057	0.056	0.050	−20	4	0.054	0.051	0.047
	3	0.057	0.052	0.049		8	0.056	0.051	0.050		12	0.057	0.051	0.048
	4	0.063	0.054	0.049		12	0.054	0.052	0.046		36	0.050	0.050	0.048
	5	0.055	0.052	0.052		20	0.054	0.047	0.046		60	0.052	0.049	0.043
−50	2	0.050	0.053	0.059	−50	4	0.052	0.050	0.050	−50	4	0.052	0.052	0.051
	3	0.051	0.055	0.051		8	0.050	0.051	0.050		12	0.048	0.052	0.051
	4	0.050	0.053	0.051		12	0.049	0.048	0.052		36	0.052	0.053	0.047
	5	0.051	0.051	0.053		20	0.051	0.050	0.049		60	0.053	0.050	0.046

Table 11
Long−horizon univariate predictive regressions−Monthly data

This table presents the results of long-horizon univariate predictive regression models, as in equation (30), during the sample periods January 1927−December 2012 (Panel A) and January 1952−December 2012 (Panel B), for various horizons (*K*-mths). The dependent variable is the cumulative S&P 500 value-weighted log excess return from month *t* to month *t+K-1*, corresponding to a horizon of *K* months, and the lagged persistent regressor is each of the following variables defined in Section 3: Dividend payout ratio (d/e), long-term yield (lty), dividend yield (d/y), dividend price ratio (d/p), T-bill rate (tbl), earnings price ratio (e/p), book-to-market value ratio (b/m), default yield spread (dfy), net equity expansion (ntis), term spread (tms) and inflation rate (inf). The table reports the long-horizon Wald statistic, defined in equation (34), under the null hypothesis that the slope coefficient of the long-horizon univariate predictive regression estimated via the proposed instrumental variable (IVX) approach, is equal to zero (i.e., no predictability). *, ** and *** imply rejection of the null hypothesis at 10%, 5% and 1% level respectively.

Panel A: January 1927−December 2012

K-mths	d/e	lty	d/y	d/p	tbl	e/p	b/m	dfy	ntis	tms	inf
4	0.138	0.752	2.322	2.271	1.413	3.978**	4.851**	0.054	4.805**	1.125	0.781
12	0.005	0.195	3.492*	3.230*	0.947	4.538**	5.767**	0.124	9.123***	2.156	0.528
24	0.472	0.061	3.772*	3.782*	0.774	3.335*	4.501**	0.141	8.784***	3.080*	0.022
36	0.803	0.039	3.415*	3.452*	0.918	2.806*	3.866**	0.105	6.816***	5.025**	0.001
48	0.422	0.021	3.150*	3.234*	0.668	3.418*	3.788*	0.222	4.960**	4.642**	0.053
60	0.637	0.024	2.912*	3.018*	0.525	3.044*	2.970*	0.232	4.309**	4.022**	0.057

Panel B: January 1952−December 2012

K-mths	d/e	lty	d/y	d/p	tbl	e/p	b/m	dfy	ntis	tms	inf
4	1.522	0.821	1.517	1.386	2.483	0.372	0.367	0.866	0.006	3.367*	5.507**
12	1.717	0.133	1.810	1.763	1.406	0.761	0.642	0.549	0.005	4.422**	8.328***
24	4.392**	0.009	1.584	1.639	0.651	0.286	0.241	0.048	0.147	3.494*	3.670*
36	5.779**	0.000	1.269	1.306	0.449	0.203	0.063	0.014	0.119	3.654*	2.400
48	3.317*	0.045	0.901	0.932	0.157	0.467	0.050	0.010	0.040	3.388*	2.297
60	3.856**	0.127	0.883	0.896	0.039	0.541	0.112	0.093	0.001	3.412*	1.311

Table 12
Long−horizon univariate predictive regressions−Quarterly data

This table presents the results of long-horizon univariate predictive regression models, as in equation (30), during the sample periods 1927Q1−2012Q4 (Panel A) and 1952Q1−2012Q4 (Panel B), for various horizons (*K*-qtrs). The dependent variable is the cumulative S&P 500 value-weighted log excess return from quarter *t* to quarter *t+K-1*, corresponding to a horizon of *K* quarters, and the lagged persistent regressor is each of the following variables defined in Section 3: Dividend payout ratio (d/e), long-term yield (lty), dividend yield (d/y), dividend price ratio (d/p), T-bill rate (tbl), earnings price ratio (e/p), book-to-market value ratio (b/m), default yield spread (dfy), net equity expansion (ntis), term spread (tms), inflation rate (inf) and consumption-wealth ratio (cay). The table reports the long-horizon Wald statistic, defined in equation (34), under the null hypothesis that the slope coefficient of the long-horizon univariate predictive regression estimated via the proposed instrumental variable (IVX) approach, is equal to zero (i.e., no predictability). *, ** and *** imply rejection of the null hypothesis at 10%, 5% and 1% level respectively.

Panel A: 1927Q1−2012Q4

K-qtrs	d/e	lty	d/y	d/p	tbl	e/p	b/m	dfy	ntis	Tms	inf
4	0.000	0.173	3.537*	3.362*	0.746	4.221**	5.750**	0.139	7.672***	1.564	0.116
8	0.424	0.047	3.567*	3.648*	0.614	3.010*	4.207**	0.170	6.135**	2.475	0.021
12	0.703	0.029	3.190*	3.233*	0.697	2.461	3.428*	0.112	4.466**	3.827*	0.036
16	0.378	0.017	2.771*	2.954*	0.510	2.906*	3.181*	0.201	3.063*	3.496*	0.083
20	0.527	0.017	2.562	2.744*	0.408	2.623	2.506	0.203	2.419	3.158*	0.061

Panel B: 1952Q1−2012Q4

K-qtrs	d/e	lty	d/y	d/p	tbl	e/p	b/m	dfy	ntis	Tms	inf	Cay
4	1.409	0.132	1.902	1.902	1.201	0.857	0.824	0.391	0.022	3.569*	5.511**	11.022***
8	3.516*	0.005	1.524	1.686	0.530	0.361	0.352	0.030	0.088	2.977*	2.585	8.794***
12	4.865**	0.000	1.269	1.348	0.353	0.244	0.113	0.003	0.079	2.993*	1.784	7.326***
16	2.960*	0.034	0.895	0.961	0.135	0.463	0.062	0.007	0.038	2.809*	1.693	6.048**
20	3.247*	0.112	0.878	0.911	0.034	0.558	0.126	0.069	0.002	2.974*	0.927	5.000**

Table 13
Long−horizon predictive regressions with multiple regressors

This table presents the results of long-horizon predictive regression models with multiple regressors, as in equation (30). Panel A contains the results for monthly data and Panel B contains the results for quarterly data. Each panel reports results for the full sample period, 1927−2012, and the subperiod 1952−2012. Results are reported for various horizons (*K*-mths in Panel A and *K*-qtrs in Panel B). In Panel A, the dependent variable is the cumulative S&P 500 value-weighted log excess return from month *t* to month *t+K-1*, corresponding to a horizon of *K* months. In Panel B, the dependent variable is the cumulative S&P 500 value-weighted log excess return from quarter *t* to quarter *t+K-1*, corresponding to a horizon of *K* quarters. The lagged persistent regressors are combinations of the following variables: Earnings price ratio (e/p), T-bill rate (tbl), default yield spread (dfy), net equity expansion (ntis) and consumption-wealth ratio (cay). The combination of regressors used in each presented case is the one derived from the general-to-specific approach for 1-period regressions, as described in Section 4.2 and presented in Tables 8 and 9. The table reports the long-horizon Wald statistic, defined in equation (34), testing the individual significance of each regressor, i.e., under the null hypothesis that the corresponding slope coefficient of the long-horizon regression estimated via the proposed instrumental variable (IVX) approach, is equal to zero. It also reports the corresponding Joint Wald statistic testing the joint significance of the employed regressors, i.e., under the null hypothesis that all slope coefficients of the long-horizon regression are jointly equal to zero. *, ** and *** imply rejection of the null hypothesis at 10%, 5% and 1% level respectively.

Panel A: Monthly data

	Period: January 1927−December 2012			Period: January 1952−December 2012		
K-mths	e/p	tbl	Joint Wald	e/p	tbl	Joint Wald
4	5.778**	3.894**	7.638**	3.257*	5.666**	5.734*
12	6.383**	3.166*	7.614**	4.093**	4.986**	5.289*
24	4.990**	2.124	5.794*	2.273	2.411	2.596
36	4.599**	1.915	5.383*	2.049	1.885	2.116
48	4.983**	1.441	5.660*	2.207	1.702	2.216
60	4.321**	1.039	4.822*	1.814	1.258	1.825

Panel B: Quarterly data

	Period: 1927Q1−2012Q4				Period: 1952Q1−2012Q4				
K-qtrs	e/p	tbl	ntis	Joint Wald	e/p	tbl	dfy	cay	Joint Wald
4	4.862**	2.922*	4.928**	13.530***	3.741*	6.114**	4.890**	16.786***	23.548***
8	3.500*	2.157	3.988**	10.393**	2.013	3.662*	1.009	11.809***	16.118***
12	3.791*	2.067	2.421	8.296**	1.477	2.683	0.062	6.733***	13.292***
16	4.150**	1.689	1.175	7.300*	1.380	1.854	0.036	3.852**	11.569**
20	3.854**	1.383	0.600	6.102	0.859	0.948	0.045	1.604	10.664**

Figure 1

Power plots for sample size n=250 and residuals' correlation coefficient $\delta = -0.95$

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of *A* increases. The solid curve ($Wald_{0.05}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line). The dashed curve ($CYQ_{0.05}$) illustrates the rejection rate using the 95% confidence interval of the Campbell and Yogo (2006) *Q*-test. The dash-dot curve ($JM_{0.05}$) illustrates the rejection rate using the $\pi^*_{0.05}$ statistic of Jansson and Moreira (2006). Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 with 10,000 repetitions for a sample size of n=250, correlation coefficient between the residuals of regressions (22) and (23) $\delta = -0.95$ and no autocorrelation in the residuals of the autoregressive equation, i.e., $\phi = 0$ in (24).

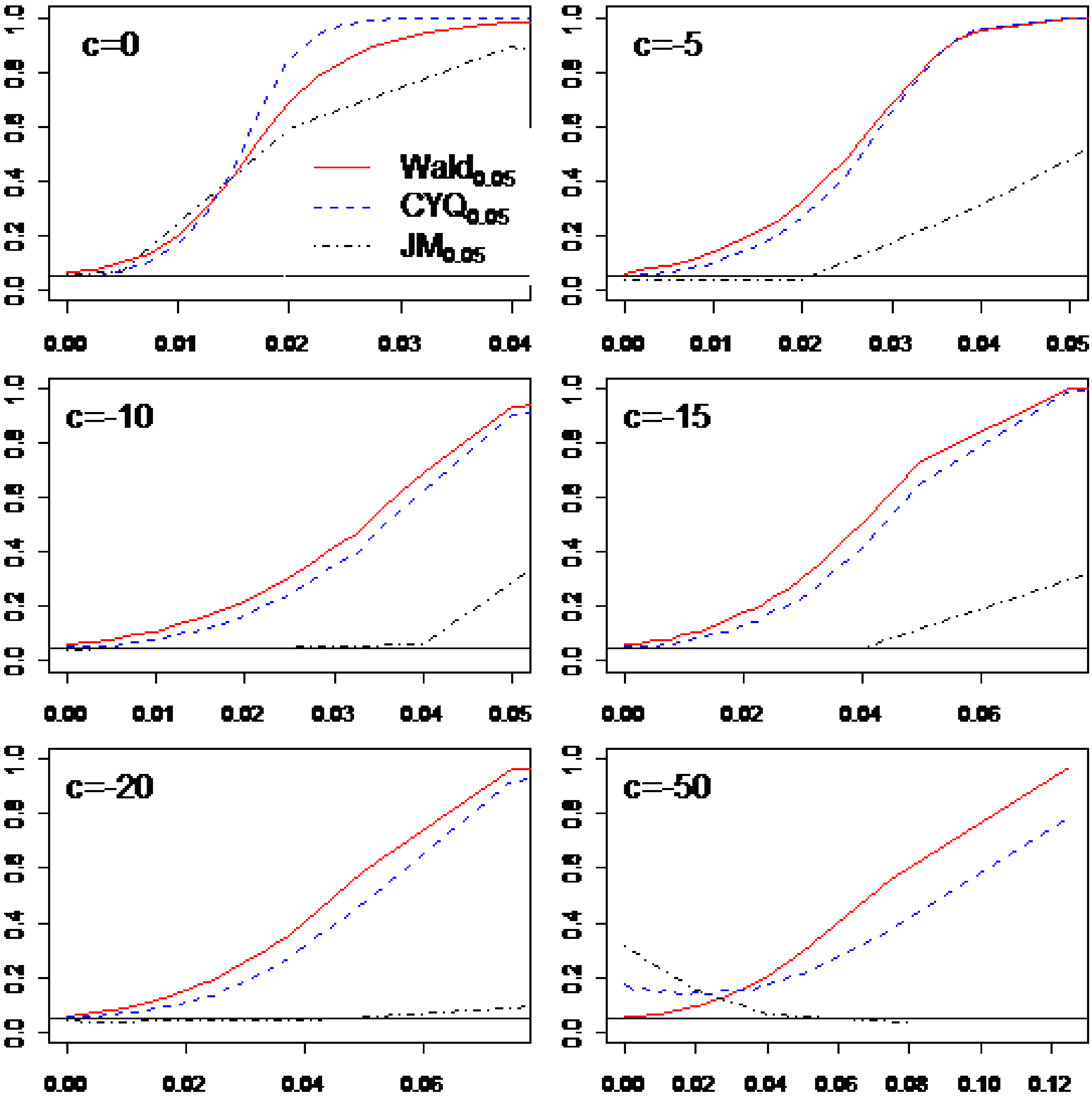


Figure 2

Power plots for sample size n=250 and residuals' correlation coefficient $\delta = -0.5$

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of *A* increases. The solid curve ($Wald_{0.05}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line). The dashed curve ($CYQ_{0.05}$) illustrates the rejection rate using the 95% confidence interval of the Campbell and Yogo (2006) *Q*-test. The dash-dot curve ($JM_{0.05}$) illustrates the rejection rate using the $\pi^*_{0.05}$ statistic of Jansson and Moreira (2006). Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 with 10,000 repetitions for a sample size of n=250, correlation coefficient between the residuals of regressions (22) and (23) $\delta = -0.5$ and no autocorrelation in the residuals of the autoregressive equation, i.e., $\phi = 0$ in (24).

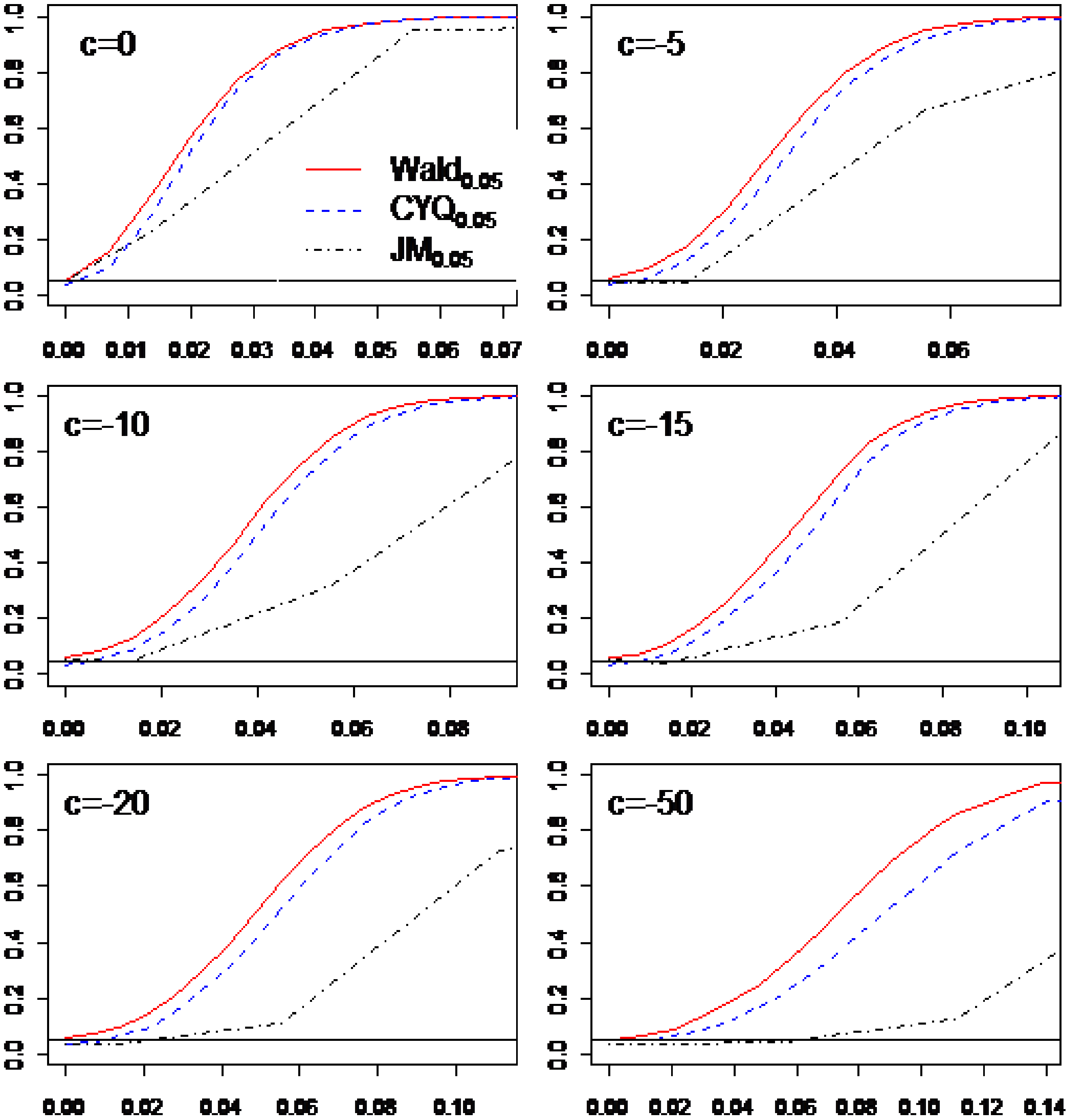


Figure 3

Power plots for sample size n=250 and residuals' correlation coefficient $\delta = 0$

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of *A* increases. The solid curve ($\text{Wald}_{0.05}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line). The dashed curve ($\text{CYQ}_{0.05}$) illustrates the rejection rate using the 95% confidence interval of the Campbell and Yogo (2006) *Q*-test. The dash-dot curve ($\text{JM}_{0.05}$) illustrates the rejection rate using the $\pi^*_{0.05}$ statistic of Jansson and Moreira (2006). Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 with 10,000 repetitions for a sample size of n=250, correlation coefficient between the residuals of regressions (22) and (23) $\delta = 0$ and no autocorrelation in the residuals of the autoregressive equation, i.e., $\phi = 0$ in (24).

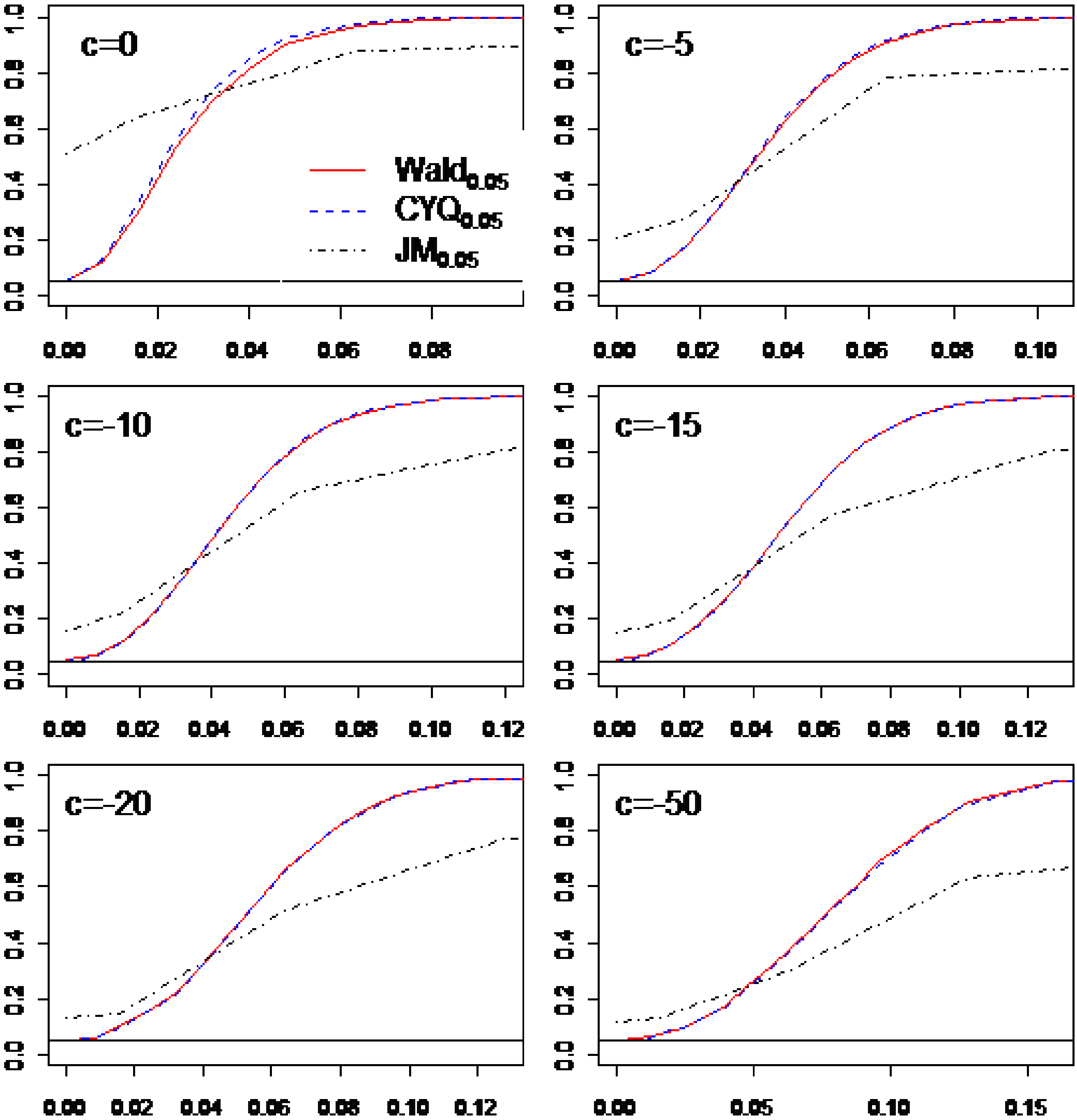


Figure 4

Power plots for joint Wald test with multiple regressors (Correlation set 1)

This figure shows the rejection rates for the joint Wald test defined in (19), with 5% nominal size, under the null hypothesis $H_0 : A = 0_{1x3}$, i.e., that all three coefficients in vector *A* are equal to zero, as the true value of each regressor coefficient A_i increases. The joint Wald test is based on the multivariate predictive system in (26), with three regressors exhibiting different degrees of persistence (unit root, local-to-unity and stationary). The solid curve (Wald^{UR}) illustrates the rejection rate for the joint Wald test as the true value of the unit root regressor coefficient increases. The dashed curve (Wald^{LTU}) illustrates the corresponding rejection rate as the true value of the local-to-unity regressor coefficient increases. The dotted curve ($\text{Wald}^{Stationary}$) illustrates the corresponding rejection rate as the true value of the stationary regressor coefficient increases. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.4 with 10,000 repetitions for different sample sizes: n=100, 250, 500 and 1,000. The correlation coefficients (δ's) between the residuals of regressions (26) and (27) are estimated using S&P 500 value-weighted log excess return (regressand), earnings-price ratio (UR), T-bill rate (LTU) and inflation rate (Stationary) with *monthly* data for the period 1927−2012, i.e., Correlation Set 1. The utilized autocorrelation coefficients (φ's) for the autoregressions are the corresponding sample estimates for each of the three regressors mentioned above.

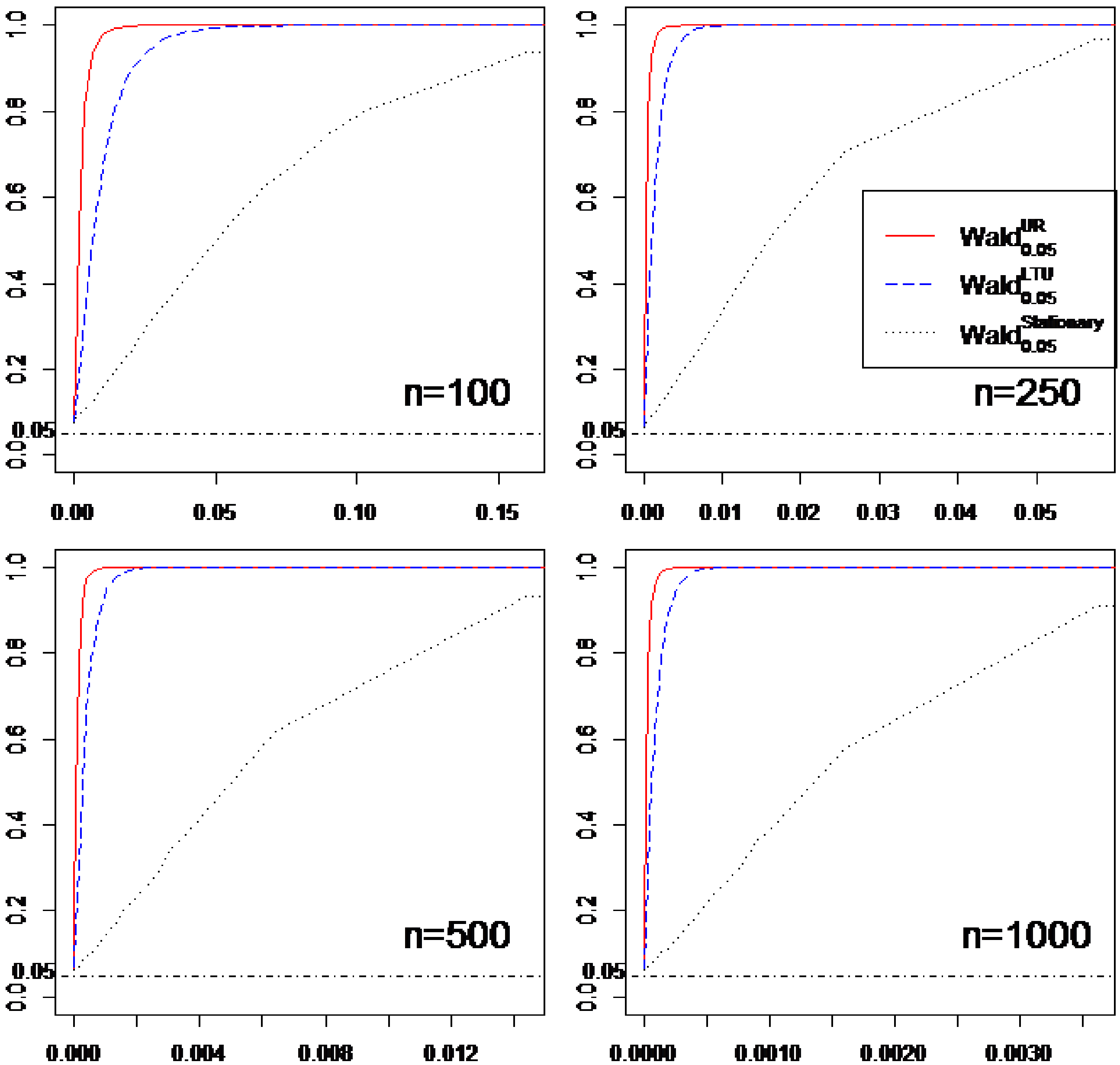
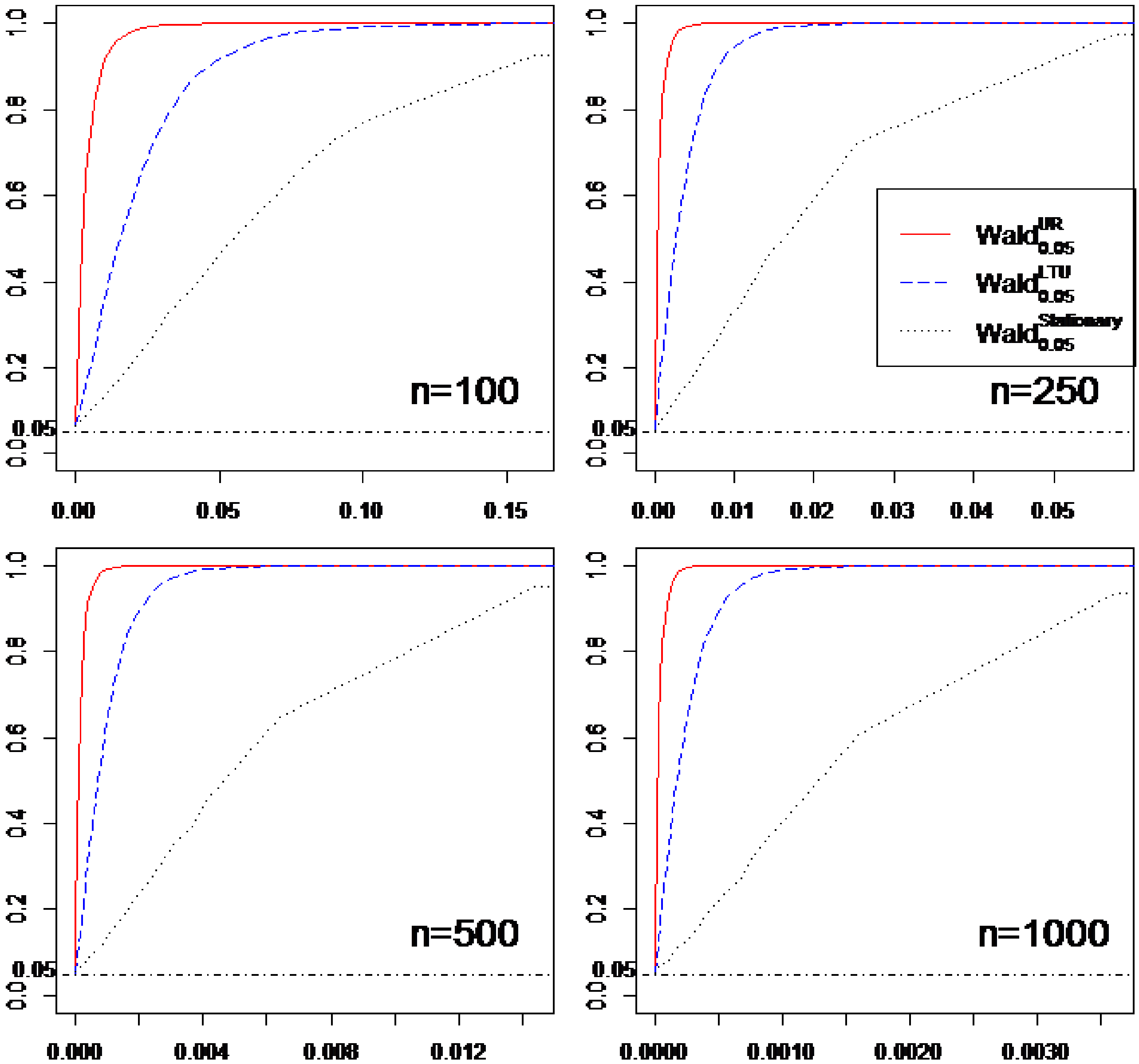


Figure 5

Power plots for joint Wald test with multiple regressors (Correlation set 2)

This figure shows the rejection rates for the joint Wald test defined in (19), with 5% nominal size, under the null hypothesis $H_0 : A = 0_{1x3}$, i.e., that all three coefficients in vector *A* are equal to zero, as the true value of each regressor coefficient A_i increases. The joint Wald test is based on the multivariate predictive system in (26), with three regressors exhibiting different degrees of persistence (unit root, local-to-unity and stationary). The solid curve ($Wald^{UR}$) illustrates the rejection rate for the joint Wald test as the true value of the unit root regressor coefficient increases. The dashed curve ($Wald^{LTU}$) illustrates the corresponding rejection rate as the true value of the local-to-unity regressor coefficient increases. The dotted curve ($Wald^{Stationary}$) illustrates the corresponding rejection rate as the true value of the stationary regressor coefficient increases. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.4 with 10,000 repetitions for different sample sizes: n=100, 250, 500 and 1,000. The correlation coefficients (δ's) between the residuals of regressions (26) and (27) are estimated using S&P 500 value-weighted log excess return (regressand), earnings-price ratio (UR), T-bill rate (LTU) and inflation rate (Stationary) with *quarterly* data for the period 1927−2012, i.e., Correlation Set 2. The utilized autocorrelation coefficients (φ's) for the autoregressions are the corresponding sample estimates for each of the three regressors mentioned above.



Robust Econometric Inference for Stock Return Predictability
Online Appendix

Alexandros Kostakis, Tassos Magdalinos and Michalis P. Stamatogiannis

1. Proofs of Theorem 1 and Theorem A

This section establishes the two main asymptotic results of the paper: the mixed Gaussianity property of the IVX estimator (Theorem A) and the chi–squared limit distribution of the IVX–Wald test statistic (Theorem 1) under Assumption INNOV. The proofs employ some useful auxiliary results that are established independently. In particular, new limit theory for IVX estimation and inference is established in the presence of conditional heteroskedasticity in the innovation errors of the general form presented in Assumption INNOV(ii) of the paper.

1.1 Introduction

We consider the system of predictive regressions

$$y_t = \mu + A x_{t-1} + \varepsilon_t, \tag{1}$$

$$x_t = R_n x_{t-1} + u_t, \tag{2}$$

$$R_n = I_r + \frac{C}{n^\alpha} \quad \text{for some } \alpha \geq 0, \tag{3}$$

with innovations ε_t, u_t satisfying Assumption INNOV and the IVX instrument process $\tilde{z}_t = R_{nz}\tilde{z}_{t-1} + \Delta x_t$ based on the matrix

$$R_{nz} = I_r + \frac{C_z}{n^\beta}, \quad \beta \in (0,1), \quad C_z < 0 \tag{4}$$

for given values of β and C_z.

1.2 Auxiliary results

We begin by establishing two auxiliary results that facilitate the proof of Theorem A and Theorem 1. The first result characterizes the asymptotic behavior of the sample mean of the IVX instruments $\tilde{z}_t$. We employ the shorthand notation $a \wedge b = \min(a,b)$ and $a \vee b = \max(a,b)$ throughout. For a given matrix M, $\|M\|$ denotes the spectral norm (the square root of the maximal eigenvalue of the matrix $M'M$) and $\|M\|_{L_p}$ denotes the usual L_p norm of vec(M). For brevity, we refer to Phillips and Magdalinos (2009) as PM and to Magdalinos and Phillips (2009) as MP.

Lemma B1. *Let α and β be defined by (3) and (4) respectively. The following approximations are valid as $n \to \infty$:*

(i) *When* $0 < \beta < \alpha$:

$$n^{-\left(\frac{1\wedge\alpha}{2}+\beta\right)} \sum_{t=1}^{n} \tilde{z}_{t-1} = -C_z^{-1} n^{-(1\wedge\alpha)/2} x_n + O_p\left(n^{-\frac{(1\wedge\alpha)-\beta}{2}}\right).$$

(ii) *When* $0 \le \alpha \le \beta < 1$: $\sum_{t=1}^{n} \tilde{z}_t = O_p\left(n^{\alpha+\frac{\beta}{2}}\right)$.

Proof. By Proposition A2 of PM,

$$\sup_{1\le t\le n} E\left\|\psi_{nt}\right\|^2 = O\left(n^{[(1\wedge\alpha)\vee\beta]+2(\alpha\wedge\beta)}\right) \quad \text{for all } \alpha \ge 0,\ \beta \in (0,1)\,. \tag{5}$$

Summing the recursion $\tilde{z}_t = R_{nz}\tilde{z}_{t-1} + \Delta x_t$ over $\{1,...,n\}$ yields

$$x_n - x_0 = \sum_{t=1}^{n} \tilde{z}_t - R_{nz}\sum_{t=1}^{n} \tilde{z}_{t-1} = \tilde{z}_n + (I_r - R_{nz})\sum_{t=1}^{n} \tilde{z}_{t-1}$$

or equivalently, since $R_{nz} - I_r = C_z/n^{\beta}$,

$$\sum_{t=1}^{n} \tilde{z}_{t-1} = C_z^{-1} n^{\beta}\left(\tilde{z}_n - x_n + x_0\right). \tag{6}$$

For part (i), using the decomposition $\tilde{z}_t = z_t + \frac{C}{n^\alpha}\psi_{nt}$ and (5) with $\beta < \alpha$ we obtain that

$$\tilde{z}_n = z_n + \frac{C}{n^\alpha}\psi_{nn} = O_p\left(n^{\beta/2}\right) + O_p\left(\frac{n^\beta}{n^{\alpha/2}}\right) = O_p\left(n^{\beta/2}\right)$$

since $n^{\beta-\alpha/2} = n^{\beta/2} n^{-(\alpha-\beta)/2} = o\left(n^{\beta/2}\right)$. Therefore, (6) implies that

$$\frac{1}{n^{\frac{\alpha}{2}+\beta}} \sum_{t=1}^{n} \tilde{z}_{t-1} = -C_z^{-1} \frac{1}{n^{\alpha/2}} x_n + O_p\left(n^{-(\alpha-\beta)/2}\right)$$

as required. For part (ii), applying the decomposition

$$\tilde{z}_t = x_t - R_{nz}^t x_0 + \frac{C_z}{n^\beta} \psi_{nt} \tag{7}$$

to (6) we obtain

$$\tilde{z}_n - x_n = \frac{C_z}{n^\beta} \psi_{nn} - R_{nz}^n x_0 = O_p\left(\frac{n^\alpha}{n^{\beta/2}}\right) \tag{8}$$

since (5) implies that $\psi_{nn} = O_p\left(n^{\alpha+\beta/2}\right)$ when $0 \leq \alpha \leq \beta$. Now part (ii) follows by combining (6) and (8).

The second auxiliary result characterizes IVX limit theory for stationary regressors. In accordance to notation of the paper, we denote the undemeaned regression matrices by

$$Y = \left(y_1', ..., y_n'\right)', \quad X = \left(x_0', ..., x_{n-1}'\right)' \quad \text{and} \quad \mathcal{E} = \left(\varepsilon_1', ..., \varepsilon_n'\right)'$$

and the demeaned regression matrices by

$$\begin{aligned} \underline{Y} &= \left(y_1' - \bar{y}_n', ..., y_n' - \bar{y}_n'\right)', \\ \underline{X} &= \left(x_0' - \bar{x}_{n-1}', ..., x_{n-1}' - \bar{x}_{n-1}'\right)' \\ \underline{\mathcal{E}} &= \left(\varepsilon_1' - \bar{\varepsilon}_n', ..., \varepsilon_n' - \bar{\varepsilon}_n'\right)'. \end{aligned}$$

Lemma B2. *Let x_t be a stable-root regressor belonging to the persistence class P(iv) with $C < 0$ and $\alpha = 0$ and let*

$$x_{0,t} = \sum_{j=0}^{\infty} R^j u_{t-j}, \quad R = I_r + C \tag{9}$$

be a stationary version of x_t. The following approximations are valid for any $\beta \in (0,1)$ under Assumption INNOV as $n \to \infty$:

(i) $n^{-1} \underline{X}' \tilde{Z} = n^{-1} X' X + o_p(1)$

(ii) $n^{-1} \tilde{Z}' \tilde{Z} = n^{-1} X' X + o_p(1)$

(iii) $n^{-1}X'X = n^{-1}\sum_{t=1}^{n} x_{0,t-1}x_{0,t-1}' + o_p(1)$

(iv) $n^{-1/2}\underline{\mathcal{E}}'\tilde{Z} = n^{-1/2}\mathcal{E}'X + o_p(1) = n^{-1/2}\sum_{t=1}^{n} \varepsilon_t x_{0,t-1}' + o_p(1)$.

Proof. By assumption $0 = \alpha < \beta < 1$, so Lemma B1(ii) implies that $\sum_{t=1}^{n} \tilde{z}_t = O_p\left(n^{\beta/2}\right)$. For part (i), applying the decomposition (7) and using the fact that $\bar{x}_{n-1} = O_p\left(n^{-1/2}\right)$ we obtain

$$\begin{aligned} n^{-1}\underline{X}'\tilde{Z} &= n^{-1}\sum_{t=1}^{n} x_{t-1}\tilde{z}_{t-1}' - \bar{x}_{n-1}\bar{z}_{n-1}' \\ &= \frac{1}{n}\sum_{t=1}^{n} x_{t-1}x_{t-1}' + \frac{1}{n^{1+\beta}}\sum_{t=1}^{n} x_{t-1}\psi_{nt-1}'C_z + O_p\left(\frac{1}{n^{1-\beta}}\right) \\ &= \frac{1}{n}X'X + O_p\left(\frac{1}{n^{\beta/2}}\right) + O_p\left(\frac{1}{n^{1-\beta}}\right) \end{aligned}$$

because the Cauchy–Schwarz inequality yields

$$\begin{aligned} \left\| \frac{1}{n^{1+\beta}}\sum_{t=1}^{n} x_{t-1}\psi_{nt-1}' \right\|_{L_1} &\leq \frac{1}{n^{1+\beta}}\sum_{t=1}^{n} E\left(\|x_{t-1}\|\,\|\psi_{nt-1}\|\right) \\ &\leq \frac{\left(E\|x_1\|^2\right)^{1/2}}{n^{1+\beta}}\sum_{t=1}^{n}\left(E\|\psi_{nt-1}\|^2\right)^{1/2} \\ &\leq \left(E\|x_1\|^2\right)^{1/2}\left(\sup_{1\leq t\leq n} E\|\psi_{nt}\|^2\right)^{1/2}\frac{1}{n^{\beta}} = O\left(\frac{1}{n^{\beta/2}}\right) \end{aligned}$$

since (5) with $\alpha = 0$ implies that $\sup_{1\leq t\leq n} E\|\psi_{nt}\|^2 = O\left(n^{\beta}\right)$. For part (ii), (7) implies the identity

$$\begin{aligned} \frac{1}{n}\tilde{Z}'\tilde{Z} &= \frac{1}{n}X'X + \frac{1}{n^{1+\beta}}\sum_{t=1}^{n} x_{t-1}\psi_{nt-1}'C_z + \frac{1}{n^{1+\beta}}C_z\sum_{t=1}^{n}\psi_{nt-1}x_{t-1}' \\ &\quad + \frac{1}{n^{1+2\beta}}C_z\sum_{t=1}^{n}\psi_{nt-1}\psi_{nt-1}'C_z + O_p\left(\frac{1}{n^{1-\beta}}\right), \end{aligned}$$

where all terms containing the initial condition x_0 are included in the remainder. From part (i), we know that the second and third terms on the right have order $O_p\left(n^{-\beta/2}\right)$. For the last term on the right hand side, (5) implies that

$$\left\| \frac{1}{n^{1+2\beta}}\sum_{t=1}^{n}\psi_{nt-1}\psi_{nt-1}' \right\|_{L_1} \leq \frac{1}{n^{1+2\beta}}\sum_{t=1}^{n} E\|\psi_{nt-1}\|^2 = O\left(\frac{1}{n^{\beta}}\right),$$

establishing part (ii). For part (iii), (7) again yields

$$\begin{aligned} n^{-1/2}\underline{\mathcal{E}}'\tilde{Z} &= n^{-1/2}\sum_{t=1}^{n}\varepsilon_t\tilde{z}_{t-1}' - n^{1/2}\bar{\varepsilon}_n\bar{z}_{n-1}' \\ &= n^{-1/2}\sum_{t=1}^{n}\varepsilon_t x_{t-1}' + \frac{1}{n^{\frac{1}{2}+\beta}}\sum_{t=1}^{n}\varepsilon_t\psi_{nt-1}'C_z + O_p\left(\frac{1}{n^{\frac{1-\beta}{2}}}\right) \end{aligned}$$

and the result follows by (5) since the second term on the right is a martingale array satisfying

$$E\left\|\frac{1}{n^{\frac{1}{2}+\beta}}\sum_{t=1}^{n}\left(\psi_{nt-1}\otimes\varepsilon_t\right)\right\|^2 = \frac{E\left\|\varepsilon_1\right\|^2}{n^{1+2\beta}}\sum_{t=1}^{n}E\left\|\psi_{nt-1}\right\|^2 \leq \frac{E\left\|\varepsilon_1\right\|^2}{n^{2\beta}}\sup_{t\leq n}E\left\|\psi_{nt}\right\|^2 = O\left(\frac{1}{n^{\beta}}\right).$$

Note that the above approximations employ unconditional moment bounds and hence apply under both part (i) and part (ii) of Assumption INNOV. To show the second asymptotic equivalence of part (iv), the identity $x_{t-1} = R^{t-1}x_0 + \sum_{j=1}^{t-1}R^{j-1}u_{t-j}$ yields

$$x_{0,t-1} - x_{t-1} = \sum_{j=t}^{\infty}R^{j-1}u_{t-j} - R^{t-1}x_0. \tag{10}$$

Using (10), we obtain

$$\begin{aligned} n^{-1/2}\left\|\mathcal{E}'X - \sum_{t=1}^{n}\varepsilon_t x_{0,t-1}'\right\| &= n^{-1/2}\left\|\sum_{t=1}^{n}\varepsilon_t\left(x_{0,t-1} - x_{t-1}\right)'\right\| \\ &\leq n^{-1/2}\sum_{t=1}^{n}\sum_{j=t}^{\infty}\left\|R\right\|^{j-1}\left\|\varepsilon_t\right\|\left\|u_{t-j}\right\| + n^{-1/2}\left\|x_0\right\|\sum_{t=1}^{n}\left\|R\right\|^{t-1}\left\|\varepsilon_t\right\| \\ &= O_p\left(n^{-1/2}\left\|x_0\right\|\right) \end{aligned}$$

as $\sum_{t=1}^{n}\left\|R\right\|^{t-1}\left\|\varepsilon_1\right\|_{L_1} \leq \left\|\varepsilon_1\right\|_{L_2}\sum_{t=0}^{\infty}\left\|R\right\|^t$ and the first term is bounded in L_1 norm by

$$n^{-1/2}\left\|\varepsilon_1\right\|_{L_2}\left\|u_1\right\|_{L_2}\sum_{t=1}^{n}\sum_{j=t}^{\infty}\left\|R\right\|^{j-1} \leq n^{-1/2}\left\|\varepsilon_1\right\|_{L_2}\left\|u_1\right\|_{L_2}\left(\sum_{j=0}^{\infty}\left\|R\right\|^j\right)^2.$$

To show part (iii), $x_{0,t-1}x_{0,t-1}' - x_{t-1}x_{t-1}' = x_{0,t-1}(x_{0,t-1} - x_{t-1})' + (x_{0,t-1} - x_{t-1})x_{t-1}'$ and (10) yield

$$\begin{aligned}
n^{-1}\left\|X'X - \sum_{t=1}^{n} x_{0,t-1}x_{0,t-1}'\right\| &\leq n^{-1}\sum_{t=1}^{n}\left(\|x_{0,t-1}\| + \|x_{t-1}\|\right)\|x_{0,t-1} - x_{t-1}\| \\
&\leq n^{-1}\sum_{t=1}^{n}\|x_{0,t-1}\|\,\|x_{0,t-1} - x_{t-1}\| \\
&\quad + n^{-1}\|x_0\|\sum_{t=1}^{n}\|x_{0,t-1} - x_{t-1}\|\,\|R\|^{t-1} \\
&\quad + n^{-1}\sum_{t=1}^{n}\sum_{j=1}^{t-1}\|R\|^{j-1}\|x_{0,t-1} - x_{t-1}\|\,\|u_{t-j}\|
\end{aligned}$$

$$\begin{aligned}
&\leq n^{-1}\sum_{t=1}^{n}\sum_{j=t}^{\infty}\|R\|^{j-1}\|x_{0,t-1}\|\,\|u_{t-j}\| + \|x_0\|\,n^{-1}\sum_{t=1}^{n}\|R\|^{t-1}\|x_{0,t-1}\| \\
&\quad + n^{-1}\|x_0\|^2\sum_{t=1}^{n}\|R\|^{2(t-1)} + n^{-1}\|x_0\|\sum_{t=1}^{n}\|R\|^{t-1}\sum_{j=t}^{\infty}\|R\|^{j-1}\|u_{t-j}\| \\
&\quad + n^{-1}\sum_{t=1}^{n}\sum_{j=1}^{t-1}\|R\|^{j-1}\sum_{l=t}^{\infty}\|R\|^{l-1}\|u_{t-l}\|\,\|u_{t-j}\| \\
&\quad + n^{-1}\|x_0\|\sum_{t=1}^{n}\|R\|^{t-1}\sum_{j=1}^{t-1}\|R\|^{j-1}\|u_{t-j}\| = O_p\left(n^{-1}\|x_0\|^2\right)
\end{aligned}$$

using the L_2 Cauchy Schwarz inequality and stationarity of $x_{0,t-1}$. This shows part (iii).

The third auxiliary result shows that the limit theory for IVX sample moments is invariant to the presence of conditional heteroskedasticity in the innovation sequence.

Lemma B3. *Under Assumption INNOV(ii) the sample moments*

$$n^{-1-\alpha}\sum_{t=1}^{n} x_{t-1}x_{t-1}',\ n^{-1-\beta}\sum_{t=1}^{n} z_{t-1}z_{t-1}',\ n^{-1-(\alpha\wedge\beta)}\sum_{t=1}^{n} x_{t-1}\tilde{z}_{t-1}',\ \text{and}\ n^{-1-(\alpha\wedge\beta)}\sum_{t=1}^{n}\tilde{z}_{t-1}\tilde{z}_{t-1}'$$

have the same limit distributions as under Assumption INNOV(i) for a regressor x_t belonging to any of the persistence classes P(i)-P(iv).

Proof. First note that Lemma 3.1 (ii), (iii) and Lemma 3.5 (ii) of PM are established using unconditional moment bounds and continue to hold under conditional heteroskedasticity.

Therefore the approximations

$$\frac{1}{n^{1+\beta}}\sum_{t=1}^{n}x_{t-1}\tilde{z}_{t-1}' = \frac{1}{n^{1+\beta}}\sum_{t=1}^{n}x_{t-1}z_{t-1}' - \frac{1}{n^{1+\alpha}}\sum_{t=1}^{n}x_{t-1}x_{t-1}'CC_z^{-1} + o_p(1), \quad \beta < \alpha \tag{11}$$

$$\frac{1}{n^{1+\alpha}}\sum_{t=1}^{n}x_{t-1}\tilde{z}_{t-1}' = \frac{1}{n^{1+\alpha}}\sum_{t=1}^{n}x_{t-1}x_{t-1}' + o_p(1), \quad \alpha < \beta \tag{12}$$

$$\frac{1}{n^{1+\beta}}\sum_{t=1}^{n}\tilde{z}_{t-1}\tilde{z}_{t-1}' = \frac{1}{n^{1+\beta}}\sum_{t=1}^{n}z_{t-1}z_{t-1}' + o_p(1), \quad \beta < \alpha \tag{13}$$

$$\frac{1}{n^{1+\alpha}}\sum_{t=1}^{n}\tilde{z}_{t-1}\tilde{z}_{t-1}' = \frac{1}{n^{1+\alpha}}\sum_{t=1}^{n}x_{t-1}x_{t-1}' + o_p(1), \quad \alpha < \beta \tag{14}$$

are valid under Assumption INNOV(ii). We need to derive the limit distributions of sample moments involving x_t separately for each persistence class.

Case 1: x_t belongs to class P(iii). In this case x_t is near stationary with autoregressive matrix $R_n = I_r + C/n^{\alpha}$ $\alpha \in (0,1)$.

For $n^{-1-\alpha}\sum_{t=1}^{n}x_{t-1}x_{t-1}'$, the identity $x_{t-1} = R_n^{t-1}x_0 + \zeta_{nt-1}$ with

$$\zeta_{nt} = \sum_{j=0}^{t-1}R_n^j u_{t-j} \tag{15}$$

and the fact that $x_0 = o_p\left(n^{\alpha/2}\right)$ imply that

$$\frac{1}{n^{1+\alpha}}\sum_{t=1}^{n}x_{t-1}x_{t-1}' = \frac{1}{n^{1+\alpha}}\sum_{t=1}^{n}\zeta_{nt-1}\zeta_{nt-1}' + o_p(1) \tag{16}$$

because $n^{-1-\alpha}\sum_{t=1}^{n}\left\|R_n^{t-1}x_0\zeta_{nt-1}'\right\| \leq n^{-1-\alpha}\left\|x_0\right\|\sum_{t=1}^{n}\left\|\zeta_{nt-1}\right\| = O_p\left(n^{-\alpha/2}x_0\right) = o_p(1)$ since $\sup_{t\geq 1}E\left\|\zeta_{nt-1}\right\| \leq \sup_{t\geq 1}\left(E\left\|\zeta_{nt-1}\right\|^2\right)^{1/2} = O\left(n^{\alpha/2}\right)$. Since $\zeta_{n,n-1} = O_p\left(n^{\alpha/2}\right)$ applying the argument of MP to the recursion $\zeta_{nt} = R_n\zeta_{nt-1} + u_t$ we obtain

$$\begin{aligned}\text{vec}\left\{\frac{1}{n^{1+\alpha}}\sum_{t=1}^{n}\zeta_{nt-1}\zeta_{nt-1}'\right\} &= \left[1 + o_p(1)\right]\left(C\otimes I + I\otimes C\right)^{-1} \\ &\quad \times\text{vec}\left\{\frac{1}{n}\sum_{t=1}^{n}\zeta_{nt-1}u_t' + \frac{1}{n}\sum_{t=1}^{n}u_t\zeta_{nt-1}' + \frac{1}{n}\sum_{t=1}^{n}u_tu_t'\right\}.\end{aligned} \tag{17}$$

Under assumption INNOV(ii), strict stationarity and ergodicity of $(e_t)_{t\in\mathbb{Z}}$ together with the summability condition $\sum_{j=0}^{\infty} j \|C_j\| < \infty$ imply that the sequences $(u_t)_{t\in\mathbb{N}}$ and $(\tilde{e}_t)_{t\in\mathbb{N}}$ with

$$\tilde{e}_t = \sum_{j=0}^{\infty} \tilde{C}_j e_{t-j}, \qquad \tilde{C}_j = \sum_{k=j+1}^{\infty} C_k \tag{18}$$

are also strictly stationary and ergodic (since $\sum_{j=0}^{\infty} j \|C_j\| < \infty$ implies $\sum_{j=0}^{\infty} \left\|\tilde{C}_j\right\| < \infty$, see Phillips and Solo (1992)). Since $\left\|n^{-1} \sum_{t=1}^{n} \zeta_{nt-1} u_t' - n^{-1} \sum_{t=1}^{n} \tilde{e}_t u_t'\right\|_{L_1} \to 0$ when $\alpha \in (0,1)$ using unconditional moment bounds (see MP), the ergodic theorem (e.g. Theorem 10.6 of Kallenberg (2002)) implies that the vectorized expression on the right side of (17) converges $a.s.$ and in L_1 to

$$E\tilde{e}_t u_t' + E u_t \tilde{e}_t' + E u_t u_t' = \Lambda_{uu} + \Lambda_{uu}' + E u_1 u_1' = \Omega_{uu}.$$

Since $(C \otimes I + I \otimes C)^{-1} \mathrm{vec}(\Omega_{uu}) = \mathrm{vec}(V_C)$, (17) implies that

$$\frac{1}{n^{1+\alpha}} \sum_{t=1}^{n} \zeta_{nt-1} \zeta_{nt-1}' \to_{L_1} V_C \tag{19}$$

and the above combined with (16) shows that $n^{-1-\alpha} \sum_{t=1}^{n} x_{t-1} x_{t-1}' \to_p V_C$.

Since z_t belongs to the class of near-stationary processes P(iii) with autoregressive matrix $R_{nz} = I_r + C_z/n^{\beta}$ for all persistence regimes, and the limit

$$n^{-1-\alpha} \sum_{t=1}^{n} x_{t-1} x_{t-1}' \to_p V_C = \int_0^{\infty} e^{rC} \Omega_{uu} e^{rC} dr \tag{20}$$

is valid for an arbitrary process in the class P(iii) with autoregressive matrix $R_n = I_r + C/n^{\alpha}$, the limit

$$n^{-1-\beta} \sum_{t=1}^{n} z_{t-1} z_{t-1}' \to_p V_{C_z} = \int_0^{\infty} e^{rC_z} \Omega_{uu} e^{rC_z} dr \tag{21}$$

follows immediately from (20).

For $n^{-1-(\alpha\wedge\beta)} \sum_{t=1}^{n} x_{t-1} \tilde{z}_{t-1}'$ with $\beta < \alpha$, equation (17) of PM yields

$$\frac{1}{n^{1+\beta}} \sum_{t=1}^{n} x_{t-1} z_{t-1}' = [I_r + o_p(1)] \frac{1}{n} \left(\sum_{t=1}^{n} x_{t-1} u_t' + \sum_{t=1}^{n} u_t z_{t-1}' + \sum_{t=1}^{n} u_t u_t' \right) \left(-C_z^{-1}\right). \tag{22}$$

Note that the asymptotic equivalence in (22) is valid for x_t belonging to any of the persistence regimes P(i)-P(iii) (the stable root case P(iv) is ruled out because $0 = \alpha < \beta$ and hence $\sum_{t=1}^{n} x_{t-1}z_{t-1}'$ does not feature in the asymptotics). When $\alpha \in (0,1)$, MP show that, independently of the conditional variance of the primitive innovations e_t, $n^{-1}\sum_{t=1}^{n} x_{t-1}u_t' = n^{-1}\sum_{t=1}^{n} \tilde{e}_t u_t' + o_p(1)$, and $n^{-1}\sum_{t=1}^{n} u_t z_{t-1}' = n^{-1}\sum_{t=1}^{n} u_t \tilde{e}_t' + o_p(1)$. The ergodic theorem applied to $n^{-1}\sum_{t=1}^{n} \tilde{e}_t u_t'$ and $n^{-1}\sum_{t=1}^{n} u_t u_t'$ implies that $n^{-1-\beta}\sum_{t=1}^{n} x_{t-1}z_{t-1}' \to_p -\Omega_{uu}C_z^{-1}$. Combining this with (11) and (20) delivers the required result when $\beta < \alpha$. When $\alpha < \beta$, the result follows by (12) and (20).

For $n^{-1-(\alpha\wedge\beta)}\sum_{t=1}^{n} \tilde{z}_{t-1}\tilde{z}_{t-1}'$ the result follows directly by (13) and (21) when $\beta < \alpha$ and by (14) and (20) when $\alpha < \beta$.

Case 2: x_t belongs to classes P(i)-P(ii). In this case $\alpha = 1$ and the asymptotic behavior of $n^{-2}\sum_{t=1}^{n} x_{t-1}x_{t-1}'$ and $n^{-1-\beta}\sum_{t=1}^{n} x_{t-1}z_{t-1}'$ is driven by a functional central limit theorem on $D_{\mathbb{R}^r}[0,1]$ of the random element

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor ns \rfloor} u_j = C(1)\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor ns \rfloor} e_j + o_p(1) \Rightarrow B(s) \tag{23}$$

where $B(s)$ is a Brownian motion with covariance matrix $\Omega_{uu} = C(1)\Sigma_{ee}C(1)'$, where the first asymptotic equivalence follows by a standard application of the Phillips and Solo (1992) BN decomposition approach. The validity of the weak convergence on $D_{\mathbb{R}^r}[0,1]$ in (23) under Assumption INNOV(ii) is guaranteed by the functional central limit theorem for stationary ergodic martingale differences, e.g. Theorem 18.3 of Billingsley (1968). Since (23) and the ergodic theorem for $n^{-1}\sum_{t=1}^{n} \tilde{e}_t u_t'$ and $n^{-1}\sum_{t=1}^{n} u_t u_t'$ continue to apply and yield the same limits under Assumption INNOV(i) and INNOV(ii), the sample moments of the lemma will have the same limit distributions.

Case 3: x_t belongs to class P(iv). In this case $\alpha = 0$ and x_t is a stable root autoregression with fixed autoregressive matrix $R = I_r + C$ with $\|R\| < 1$. Given the results of Lemma B2 above, the only sample moment of interest in this case is $n^{-1}\sum_{t=1}^{n} x_{0,t-1}x_{0,t-1}'$, with $x_{0,t}$ defined in (9). Since $\sum_{j=0}^{\infty}\|R\|^j < \infty$ and $(u_t)_{t\in\mathbb{Z}}$ is strictly stationary and ergodic, $x_{0,t}$ is a strictly stationary and ergodic process. The ergodic theorem then implies that

$$n^{-1}\sum_{t=1}^{n} x_{0,t-1}x_{0,t-1}' \to_{L_1} E\left(x_{0,1}x_{0,1}'\right) = \sum_{i,j=0}^{\infty} R^i \Gamma_u(i-j) R^j \tag{24}$$

where $\Gamma_u(k) = E\left(u_t u_{t-k}'\right)$ and $n^{-1}X'X \to_p E\left(x_{0,1}x_{0,1}'\right)$ follows by Lemma B2(iii).

Lemma B4. *Under Assumption INNOV(ii) the following limits apply as $n \to \infty$:*

(i) *under P(iii),* $n^{-(1+\alpha)/2} \sum_{t=1}^{n} (x_{t-1} \otimes \varepsilon_t) \Rightarrow N(0, V_C \otimes \Sigma_{\varepsilon\varepsilon})$

(ii) $n^{-(1+\beta)/2} \sum_{t=1}^{n} (z_{t-1} \otimes \varepsilon_t) \Rightarrow N(0, V_{C_z} \otimes \Sigma_{\varepsilon\varepsilon})$

(iii) *under P(iv),* $n^{-1/2} \sum_{t=1}^{n} (x_{t-1} \otimes \varepsilon_t) \Rightarrow N\left(0, E\left[x_{0,1} x_{0,1}' \otimes \varepsilon_2 \varepsilon_2'\right]\right)$

where $x_{0,t}$ *is defined in (9).*

Proof. For part (i), x_t is a near stationary process with $\alpha \in (0,1)$. The proof is long and technically demanding and is contained in Magdalinos (2014).

For part (ii), since z_t belongs to the class near-stationary processes P(iii) with autoregressive matrix $R_{nz} = I_r + C_z/n^{\beta}$, the limit distribution of $n^{-(1+\beta)/2} \sum_{t=1}^{n} (z_{t-1} \otimes \varepsilon_t)$ can be deduced directly by part (i) above: since for an arbitrary near stationary process x_t with autoregressive matrix $R_n = I_r + C/n^{\alpha}$, the limit distribution of $n^{-(1+\alpha)/2} \sum_{t=1}^{n} (x_{t-1} \otimes \varepsilon_t)$ is Gaussian with mean zero and covariance matrix equal to the probability limit of $n^{-1-\alpha} \sum_{t=1}^{n} x_{t-1} x_{t-1}' \otimes \Sigma_{\varepsilon\varepsilon}$, the limit distribution of $n^{-(1+\beta)/2} \sum_{t=1}^{n} (z_{t-1} \otimes \varepsilon_t)$ is Gaussian with mean zero and covariance matrix equal to the probability limit of $n^{-1-\beta} \sum_{t=1}^{n} z_{t-1} z_{t-1}' \otimes \Sigma_{\varepsilon\varepsilon}$, the latter being equal to $V_{C_z} \otimes \Sigma_{\varepsilon\varepsilon}$ by (21).

For part (iii), recalling the definition of $x_{0,t}$ in (9), Lemma B2(iv) implies that it is enough to derive the limit distribution of $n^{-1/2} \sum_{t=1}^{n} (x_{0,t-1} \otimes \varepsilon_t)$. Since $\{x_{0,t-1} \otimes \varepsilon_t : t \geq 1\}$ is a strictly stationary and ergodic martingale difference, Theorem 18.3 of Billingsley (1968) implies that

$$n^{-1/2} \sum_{t=1}^{n} (x_{0,t-1} \otimes \varepsilon_t) \Rightarrow N\left(0, E\left(x_{0,t-1} x_{0,t-1}' \otimes \varepsilon_t \varepsilon_t'\right)\right)$$

and the result follows by strict stationarity.

1.3 Proof of Theorem A

We use Lemma B1 throughout. For part (i), we start with the signal matrix:

$$\begin{aligned} \frac{1}{n^{1+\beta}} \underline{X}' \tilde{Z} &= \frac{X'\tilde{Z}}{n^{1+\beta}} - \left(\frac{1}{n^{3/2}} \sum_{t=1}^{n} x_{t-1} \right) \left(\frac{1}{n^{\frac{1}{2}+\beta}} \sum_{t=1}^{n} \tilde{z}_{t-1} \right)' \\ &= \frac{X'\tilde{Z}}{n^{1+\beta}} + \left(\frac{1}{n^{3/2}} \sum_{t=1}^{n} x_{t-1} \right) \frac{1}{n^{1/2}} x_n' C_z^{-1} + o_p(1) \end{aligned}$$

using part (i) of Lemma B1. The limit distribution of $n^{-(1+\beta)}X'\tilde{Z}$ is given by Lemma 3.1(ii) and equation (20) of PM. Note that all of the above normalized sums are bounded in probability for all $\alpha > 0$. When $\alpha = 1$ in case P(ii),

$$\begin{aligned} \frac{1}{n^{1+\beta}}\underline{X}'\tilde{Z} &\Rightarrow -\left[\Omega_{uu} + \int_0^1 J_C dJ_C'\right] C_z^{-1} + \left(\int_0^1 J_C\right) J_C(1)' C_z^{-1} \\ &= -\left(\Omega_{uu} + \int_0^1 \underline{J}_C dJ_C'\right) C_z^{-1}. \end{aligned} \tag{25}$$

In the unit root case P(i), the limit distribution of $n^{-(1+\beta)}\underline{X}'\tilde{Z}$ can be obtained by substituting $C = 0$ in (25):

$$\frac{1}{n^{1+\beta}}\underline{X}'\tilde{Z} \Rightarrow -\left[\Omega_{uu} + \int_0^1 \underline{B}_x dB_x'\right] C_z^{-1}. \tag{26}$$

In the near–stationary case, $\sum_{t=1}^n x_{t-1} = O_p\left(n^{1/2+\alpha}\right)$ and $x_n = O_p\left(n^{\alpha/2}\right)$ with $\alpha < 1$ by MP. Equation (20) of PM then yields

$$\frac{1}{n^{1+\beta}}\underline{X}'\tilde{Z} = \frac{X'\tilde{Z}}{n^{1+\beta}} + o_p(1) = -\left(\Omega_{uu} + V_C C\right) C_z^{-1} + o_p(1). \tag{27}$$

Combining (25), (26) and (27) and taking into account multiplication by $-C_z^{-1}$ yields $\tilde{\Psi}_{uu}$ of Theorem A.

Next, we show that the presence of an intercept in (1) has no effect on the asymptotic behavior of the $\underline{\mathcal{E}}'\tilde{Z}$ matrix: using part (i) of Lemma B1

$$\begin{aligned} n^{-(1+\beta)/2}\underline{\mathcal{E}}'\tilde{Z} &= n^{-(1+\beta)/2}\mathcal{E}'\tilde{Z} - \left(\frac{1}{n^{1/2}}\sum_{t=1}^n \varepsilon_t\right)\left(\frac{1}{n^{1+\beta/2}}\sum_{t=1}^n \tilde{z}_{t-1}\right)' \\ &= n^{-(1+\beta)/2}\mathcal{E}'\tilde{Z} - \frac{1}{n^{\frac{1-\beta}{2}}}\frac{1}{n^{\frac{1-(\alpha\wedge 1)}{2}}} C_z^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n \varepsilon_t\right)\frac{x_n'}{n^{\frac{\alpha\wedge 1}{2}}} \\ &= n^{-(1+\beta)/2}\sum_{t=1}^n \varepsilon_t \tilde{z}_{t-1}' + O_p\left(n^{-\frac{1-\beta}{2}}\right) \\ &= n^{-(1+\beta)/2}\sum_{t=1}^n \varepsilon_t z_{t-1}' + o_p(1) \end{aligned} \tag{28}$$

by Lemma 3.1(i) of PM. The limit distribution of $n^{-(1+\beta)/2}\mathrm{vec}\left(\underline{\mathcal{E}}'\tilde{Z}\right)$ is then given by

$$n^{-(1+\beta)/2}\sum_{t=1}^n \left(z_{t-1}\otimes\varepsilon_t\right) \Rightarrow N\left(0, V_{C_z}\otimes\Sigma_{\varepsilon\varepsilon}\right)$$

established by Lemma 3.2 of PM under INNOV(i) and Lemma B4(ii) above under INNOV(ii). Lemma 3.2 of PM also establishes the asymptotic independence as $n \to \infty$ between $n^{-(1+\beta)/2}\mathcal{E}'\tilde{Z}$ and $n^{-(1+\beta)}\underline{X}'\tilde{Z}$. This completes the proof of part (i) of Theorem A.

For parts (ii)–(iv), $0 \le \alpha \le \beta < 1$. We first show that the presence of an intercept in (1) has no effect on IVX limit theory. Using the fact that $\sum_{t=1}^{n} x_{t-1} = O_p\left(n^{\frac{1}{2}+\alpha}\right)$, the signal matrix can be written as

$$\begin{aligned} n^{-(1+\alpha)}\underline{X}'\tilde{Z} &= n^{-(1+\alpha)}X'\tilde{Z} - \left(\frac{1}{n^{\frac{1}{2}+\alpha}}\sum_{t=1}^{n} x_{t-1}\right)\left(\frac{1}{n^{\frac{3}{2}}}\sum_{t=1}^{n}\tilde{z}_{t-1}\right)' \\ &= n^{-(1+\alpha)}X'\tilde{Z} + O_p\left(n^{-(1-\alpha)}n^{-(1-\beta)/2}\right). \end{aligned} \tag{29}$$

where the order of magnitude of $\sum_{t=1}^{n}\tilde{z}_{t-1}$ follows from Lemma B1(ii). Using an identical argument

$$\begin{aligned} n^{-(1+\alpha)/2}\underline{\mathcal{E}}'\tilde{Z} &= n^{-(1+\alpha)/2}\mathcal{E}'\tilde{Z} - \left(\frac{1}{n^{1/2}}\sum_{t=1}^{n}\varepsilon_t\right)\left(\frac{1}{n^{1+\alpha/2}}\sum_{t=1}^{n}\tilde{z}_t\right)' \\ &= n^{-(1+\alpha)/2}\mathcal{E}'\tilde{Z} + O_p\left(n^{-(1-\alpha)/2}n^{-(1-\beta)/2}\right), \end{aligned} \tag{30}$$

so both sample moment matrices $n^{-(1+\alpha)}\underline{X}'\tilde{Z}$ and $n^{-(1+\alpha)/2}\underline{\mathcal{E}}'\tilde{Z}$ are asymptotically equivalent to $n^{-(1+\alpha)}X'\tilde{Z}$ and $n^{-(1+\alpha)}\mathcal{E}'\tilde{Z}$ respectively and the limit results of parts (ii) and (iii) of Theorem A can be deduced by Theorem 3.7 of PM.

It remains to show part (iv) of the theorem that is not included in PM. By Lemma B2,

$$\begin{aligned} \sqrt{n}\left(\tilde{A}_{IVX} - A\right) &= \frac{1}{\sqrt{n}}\underline{\mathcal{E}}'\tilde{Z}\left(\frac{1}{n}\underline{X}'\tilde{Z}\right)^{-1} \\ &= \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t x_{0,t-1}'\right)\left(\frac{1}{n}\sum_{t=1}^{n}x_{0,t-1}x_{0,t-1}'\right)^{-1} + o_p(1) \end{aligned} \tag{31}$$

with $x_{0,t}$ defined in (9). Under Assumption INNOV(ii), (24) and Lemma B4(iii) imply that $\sqrt{n}\text{vec}\left(\tilde{A}_{IVX} - A\right) \Rightarrow N(0, V)$ where

$$V = \left(\left[Ex_{0,1}x_{0,1}'\right]^{-1} \otimes I_m\right) E\left(x_{0,1}x_{0,1}' \otimes \varepsilon_2\varepsilon_2'\right)\left(\left[Ex_{0,1}x_{0,1}'\right]^{-1} \otimes I_m\right). \tag{32}$$

Positive definiteness of the moment matrix $E\left(x_{0,1}x_{0,1}'\right)$ is guaranteed by the positive definiteness of $\Sigma_{ee} = E\left(e_1 e_1'\right)$. Under Assumption INNOV(i), conditional homoskedasticity and the $2+\delta$-moment condition imposed on the martingale difference sequence e_t are sufficient for the law of large numbers (24). Also, a standard martingale central limit theorem yields

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(x_{0,t-1}\otimes\varepsilon_t\right)\Rightarrow N\left(0,E\left(x_{0,1}x_{0,1}'\right)\otimes\Sigma_{\varepsilon\varepsilon}\right)$$

giving $\sqrt{n}\mathrm{vec}\left(\tilde{A}_{IVX}-A\right)\Rightarrow N\left(0,\left[Ex_{0,1}x_{0,1}'\right]^{-1}\otimes\Sigma_{\varepsilon\varepsilon}\right)$. Note that, if the sequence ε_t is conditionally homoskedastic, the limit matrix V in (32) and $\left[Ex_{0,1}x_{0,1}'\right]^{-1}\otimes\Sigma_{\varepsilon\varepsilon}$ agree.

1.4 Proof of Theorem 1

The "undemeaned" Wald statistic

$$\begin{aligned}\tilde{W}_{IVX} &= \left(H\mathrm{vec}\tilde{A}_{IVX}-h\right)'\tilde{Q}_H^{-1}\left(H\mathrm{vec}\tilde{A}_{IVX}-h\right)\\ \tilde{Q}_H &= H\left[\left(\underline{X}'P_{\tilde{Z}}\underline{X}\right)^{-1}\otimes\hat{\Sigma}_{\varepsilon\varepsilon}\right]H'\end{aligned}$$

with $P_{\tilde{Z}}=\tilde{Z}\left(\tilde{Z}'\tilde{Z}\right)^{-1}\tilde{Z}'$ is known to satisfy $\tilde{W}_{IVX}\Rightarrow\chi^2(q)$ as $n\to\infty$ by Theorem 3.8 of PM for all $\alpha>0$, i.e., all predictors belonging to classes P(i)–P(iii) under Assumption INNOV. We need to prove that: (a) $\tilde{W}_{IVX}\Rightarrow\chi^2(q)$ for stable regressors in P(iv) with $\alpha=0$ under Assumption INNOV(i); (b) $W_{IVX}-\tilde{W}_{IVX}=o_p(1)$, i.e., establish an asymptotic equivalence between W_{IVX} in Theorem 1 and $\tilde{W}_{IVX}$ above.

For (a), letting x_t be a stable regressor in P(iv) with $\alpha=0$ and using Lemma B2 repeatedly we obtain

$$n\tilde{Q}_H = H\left[\left(\frac{X'X}{n}\right)^{-1}\otimes\hat{\Sigma}_{\varepsilon\varepsilon}\right]H'+o_p(1)\to_p Q:=H\left[\left(Ex_1x_1'\right)^{-1}\otimes\Sigma_{\varepsilon\varepsilon}\right]H'$$

by Assumption INNOV(i). Under the null hypothesis,

$$\begin{aligned}\xi_n &= \tilde{Q}_H^{-1/2}\left(H\mathrm{vec}\tilde{A}_{IVX}-h\right)=\left(n\tilde{Q}_H\right)^{-1/2}H\mathrm{vec}\sqrt{n}\left(\tilde{A}_{IVX}-A\right)\\ &= \left(n\tilde{Q}_H\right)^{-1/2}H\left[\left(\frac{X'X}{n}\right)^{-1}\otimes I_m\right]\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(x_{t-1}\otimes\varepsilon_t\right)\\ &\Rightarrow Q^{-1/2}N(0,Q)=_d N\left(0,I_q\right)\end{aligned}$$

and $\tilde{W}_{IVX} = \xi_n' \xi_n \Rightarrow \chi^2(q)$, where q is the rank of the matrix H.

Having established that $\tilde{W}_{IVX} \Rightarrow \chi^2(q)$ for all processes x_t belonging to P(i)–P(iv) under INNOV(i) and for x_t belonging to P(i)–P(iii) under INNOV(ii), Theorem 1 will follow by establishing (b). In view of the form of W_{IVX} and $\tilde{W}_{IVX}$, $W_{IVX} - \tilde{W}_{IVX} = o_p(1)$ is equivalent to

$$\left\|\mathbb{M} - \tilde{Z}'\tilde{Z} \otimes \hat{\Sigma}_{\varepsilon\varepsilon}\right\| = n\left\|\bar{z}_{n-1}\right\|^2 \left\|\hat{\Omega}_{FM}\right\| = o_p\left(\left\|\tilde{Z}'\tilde{Z}\right\|\right). \tag{33}$$

Note first that $\left\|\hat{\Omega}_{FM}\right\| = O_p(1)$ for all $\alpha \geq 0$, so we need to compare the rate of $n\left\|\bar{z}_{n-1}\right\|^2$ with that of $\left\|\tilde{Z}'\tilde{Z}\right\|$ and show that (33) is satisfied for each class P(i)–P(iv).

For $0 < \beta < \alpha$, part (i) of Lemma B1 yields

$$n\left\|\bar{z}_{n-1}\right\|^2 = \frac{1}{n}\left\|\sum_{t=1}^{n} \tilde{z}_{t-1}\right\|^2 = \frac{1}{n} O_p\left(n^{1+2\beta}\right) = O_p\left(n^{2\beta}\right) = o_p\left(\left\|\tilde{Z}'\tilde{Z}\right\|\right)$$

since $\tilde{Z}'\tilde{Z} = O_p\left(n^{1+\beta}\right)$. This establishes (33) for P(i)–P(iii) when $\beta < \alpha$.

It remains to show (33) when $0 \leq \alpha \leq \beta$. Using part (ii) of Lemma B1,

$$n\left\|\bar{z}_{n-1}\right\|^2 = \frac{1}{n}\left\|\sum_{t=1}^{n} \tilde{z}_{t-1}\right\|^2 = O_p\left(n^{2\alpha+\beta-1}\right) = o_p\left(\left\|\tilde{Z}'\tilde{Z}\right\|\right)$$

since $\left\|\tilde{Z}'\tilde{Z}\right\| = O_p\left(\left\|X'X\right\|\right) = n^{1+\alpha}$. This establishes (33), (b) and the theorem.

2. Additional Monte Carlo results

2.1 Finite–sample size for alternative values of autocorrelation coefficient ϕ

Using the univariate DGP in Section 2.1 of the main body of the study, we examine the finite–sample performance of the proposed IVX–Wald statistic and the Q–statistic of CY using alternative degrees of autocorrelation in the error term of the autoregression (ϕ). Recall that we run a 5% two–sided test under the null hypothesis $H_0 : A = 0$ for each of these two statistics. We consider sample sizes $n = 100, 250, 500, 1000$, residuals' correlation $\delta = -0.95, -0.5, 0, 0.5, 0.95$ and $C = 0, -5, -10, -20, -50$. Table A1 presents the finite–sample size of these two statistics for $\phi = 0.25$. The Wald statistic appears to have size very close to 5% in all cases considered. The Q–statistic appears to have the correct size for combinations of $\delta \in \{-0.95, 0, 0.95\}$ and

$C \in \{0, -5, -10, -20\}$. However, for $|\delta| = 0.5$ the Q–statistic becomes undersized, while for $|\delta| = 0.95$ and $C = -50$ it becomes severely oversized.

–Table A1 here–

Table A2 presents the corresponding simulation results for $\phi = -0.1$. The main conclusions regarding the comparison between the Wald statistic and the Q–statistic are very similar to the ones drawn from Table A1. The Wald statistic appears to have size very close to 5%. The Q–statistic has the correct size for $\delta = 0$, but it is severely oversized for low degrees of persistence ($C = -50$) when $|\delta| = 0.95$, while for high degrees of persistence ($-20 \leq C \leq 0$) and $|\delta| = 0.5$ it appears to be undersized. The undersizing of the Q–statistic does not disappear as the sample size increases.

–Table A2 here–

2.2 Power plots for $n = 1000$

In this subsection, we present the finite–sample power properties of the Wald, CY and JM statistics for sample size $n = 1000$. Figure A1 presents the power plots for $\delta = -0.95$, while Figures A2 and A3 present the corresponding plots for $\delta = -0.5$ and $\delta = 0$. The conclusions on the relative performance of the Wald and Q-statistic are identical to those obtained for $n = 250$ and discussed in the main body of the paper. In sum, the IVX-Wald test outperforms the Q-statistic for all persistence and correlation scenarios apart from the unit root case ($C = 0$) when $\delta \in \{-0.95, 0\}$. For $\delta = 0$, the power plots of these two statistics are almost indistinguishable. Finally, despite the increased sample size, the JM statistic still exhibits a significant lack of power relative to the other two statistics. This problem is magnified as we move away from the unit root case for all values of δ considered.

–Figures A1 to A3 here–

2.3 Power plots for autocorrelation coefficient $\phi = 0.5$

In this subsection, we present the finite–sample power properties of the Wald and the Q–statistic in the presence of autocorrelation in the error term of the autoregressive equation (ϕ). In particular, the subsequent power plots are computed using $\phi = 0.5$. Figure A4 presents the power plots for sample size $n = 250$ and $\delta = -0.95$. In the unit root case ($C = 0$), the Wald statistic is more powerful for alternatives close to the null, but for alternatives farther away, the Q–statistic becomes more powerful. For lower degrees of persistence, the Wald statistic dominates the Q–statistic. Figure A5 presents the corresponding power comparison for $n = 250$ and $\delta = -0.5$. In this case, the Wald statistic dominates the Q–statistic in terms of power for every degree of persistence considered. Finally, Figure A6 presents the corresponding power plots for $n = 250$ and $\delta = 0$. The Q–statistic is more powerful than the Wald statistic in the unit root case ($C = 0$), but for lower degrees of persistence, the two statistics appear to have almost exactly the same power.

–Figures A4–A6 here–

Figures A7, A8 and A9 present the corresponding power plots for these two statistics for $n = 1000$, $\phi = 0.5$ and residuals' correlation coefficient $\delta = -0.95$, -0.5 and 0, respectively. These plots point to the same conclusions as the ones derived from Figures A4–A6 that we discussed above.

–Figures A7–A9 here–

2.4 Alternative kernels for the estimation of the long–run covariance matrix

In this subsection, we examine the robustness of the finite–sample properties of the IVX–Wald statistic with respect to the choice of kernel for the estimation of the long–run covariance matrix.[1] In particular, apart from the Bartlett kernel that we use in the benchmark results, we alternatively use: *i*) the Parzen kernel and *ii*) the Quadratic Spectral kernel. Figure A10 illustrates the power of the IVX–Wald test for $n = 250$ and $\delta = -0.95$ using each of the aforementioned kernels. In the unit root case ($C = 0$), the Bartlett kernel appears to deliver marginally higher power, while for the rest persistence scenarios, the powers derived from these kernels appear to be almost exactly the same. Figures A11 and A12 present the corresponding power comparisons for the Wald statistic across these three kernels using $\delta = -0.5$ and $\delta = 0$, respectively. In both cases, the power plots are almost indistinguishable across the three kernels used.

–Figures A10–A12 here–

[1] We would like to thank an anonymous referee for suggesting this robustness check.

2.5 Alternative choice of lag length for the Newey–West estimator

In this subsection, we examine the robustness of the finite–sample properties of the IVX–Wald statistic when alternative lag lengths are used for the Newey–West estimator of the long–run covariance matrix.[2] In particular, apart from the truncation lag $n^{1/3}$ that we use in the benchmark results, we alternatively consider the following truncation lags: *i)* $n^{1/4}$ and *ii)* $n^{1/2}$, where n is the sample size. Figure A13 illustrates the power of the Wald statistic for $n = 250$ and $\delta = -0.95$ using each of the aforementioned lag lengths. We observe that in the unit root case ($C = 0$), the truncation lag $n^{1/2}$ appears to yield the highest power. In the rest persistence scenarios ($C < 0$), the choice of truncation lag seems to yield no difference in terms of power. This is also true regardless of the degree of regressor persistence for the case where $\delta = -0.5$, presented in Figure A14, and the case where $\delta = 0$, presented in Figure A15. In both cases, the power plots are almost indistinguishable across the three truncation lags used.

–Figures A13–A15 here–

Summarising the results presented in Figures A10–A15, the finite–sample properties of the Wald statistic are not substantially affected by the choice of the kernel or the choice of the lag length used in the estimation of the long–run covariance matrix.

2.6 Alternative values of β for the construction of instrument $\tilde{z}$

In this subsection, we examine the effect of the value of parameter β used for the construction of instruments $\tilde{z}$ on the finite–sample properties of the Wald statistic. Recalling that in our setup $\beta \in (0, 1)$, we consider 45 alternative values for $\beta \in \{0.10, 0.12, ..., 0.98\}$. The presented simulation results are derived using sample size $n = 500$. For each value of β considered, we calculate the finite–sample size of the Wald test under the null hypothesis $H_0 : A = 0$ as well as its power when the true value of A takes each of the following values: $A \in \{0.02, 0.04, 0.06\}$. As in the previous simulations, we consider various degrees of regressor persistence, i.e., $C \in \{0, -5, -10, -15, -20, -50\}$.

Figure A16 presents the rejection rates as a function of β for the case of $\delta = -0.95$. We observe that the size of the test is very close to the nominal 5% level regardless of the value of

[2] We would like to thank the Editor for suggesting this robustness check.

β. This holds true for all cases of regressor persistence considered. This observation reaffirms in finite samples the asymptotic properties of the IVX-Wald test for a very wide range of values of β. With respect to the power of the test, we find that it increases monotonically as β increases, since the rejection rate increases with β for each true value $A > 0$. A closer inspection of these power plots suggests that starting from moderate values of β, there are still considerable power gains if we further increase β towards its upper boundary, especially when the true alternative A is closer to the null. This crucial observation demonstrates that increasing the value of β offers further power gains where we need them the most, that is for true values of A close to the null.

Figures A17 and A18 present the corresponding rejection rates when $\delta = -0.5$ and $\delta = 0$, respectively. The conclusions derived from these Figures are very similar to the ones discussed above for the case of $\delta = -0.95$. Given this evidence, we can confidently argue that high values of β yield the highest level of power for the Wald test and, at the same time, yield size very close to the nominal 5%. Therefore, in the empirical implementation of our testing procedure, we use $\beta = 0.95$, which is among the highest values that β can take.

–Figures A16–A18 here–

2.7 Finite–sample properties using conditionally heteroskedastic DGP

The DGP utilized in the benchmark Monte Carlo simulations assumed homoskedasticity for the innovations of the predictive regression. In this section, we use an alternative DGP that allows these innovations to be conditionally heteroskedastic. Recalling that the asymptotic results for the proposed Wald statistic are valid under conditional heteroskedasticity too, we use a heteroskedastic DGP to examine the finite–sample properties of the statistic and compare them with the corresponding properties of the Q–statistic of CY.

In particular, we use the following GARCH(1,1) DGP for the innovations of the univariate predictive regression:

$$y_t = \mu + Ax_{t-1} + \sqrt{h_t}\varepsilon_t \qquad (34)$$

$$h_t = \omega + \alpha_1 h_{t-1} + \beta_1 \varepsilon_{t-1}^2, \qquad (35)$$

while the rest features of the DGP remain the same as in the benchmark case presented in

Section 2.1 of the main body of the study.

To render the heteroskedastic DGP empirically relevant, we estimate regression (34) using CRSP S&P 500 log excess returns as the regressand and the dividend yield as the regressor, and fit a GARCH(1,1) model to the variance of the regression residuals. Using monthly data for the full sample period, the estimated GARCH(1,1) coefficients are $\hat{\alpha}_1 = 0.13$ and $\hat{\beta}_1 = 0.85$.

Using this conditionally heteroskedastic DGP, Table A3 presents the finite–sample size of the Wald and the Q–statistic for $n = 100, 250, 500, 1000$, $\delta = -0.95, -0.5, 0, 0.5, 0.95$ and $C = 0, -5, -10, -20, -50$. We find that the Wald statistic exhibits no size distortion for all combinations considered. The Q–statistic exhibits correct size for $\delta = 0$, but it is oversized for the combination $n = 100$, $|\delta| = 0.95$ and $C = -50$, while it is undersized when $|\delta| = 0.5$.

–Table A3 here–

The introduction of conditional heteroskedasticity in the innovations of the predictive regression alters the comparative performance of the two statistics in terms of power relative to the benchmark simulation results under homoskedasticity. Figure A19 presents the power of the Wald and the Q–statistic under the conditionally heteroskedastic DGP, when $\delta = -0.95$ and $n = 1000$. In this case, the Wald statistic dominates the Q–statistic for every degree of regressor persistence considered. The same conclusion is derived from Figure A20, referring to the case where $\delta = -0.5$. However, for the case where $\delta = 0$, presented in Figure A21, we find that in the unit root case ($C = 0$), the Q–statistic has higher power than the Wald statistic. For all other degrees of regressor persistence ($C < 0$), the two statistics appear to have the same power. Very similar are the results for the other sample sizes, which are available upon request.

–Figures A19–A21 here–

2.8 Power plots for the long–horizon Wald statistic

In this subsection, we present the finite–sample power properties of the long–horizon Wald statistic for sample size $n = 1000$ and horizons $K = 12, 36, 60$ as well as for sample size $n = 500$ and horizons $K = 4, 12, 20$. Figure A22 shows the power of the long–horizon Wald statistic for $n = 1000$ and correlation $\delta = -0.95$. We find that for all horizons considered, the power of the statistic increases as the true value of A increases. Moreover, in each case, as the predictive horizon increases, the power of the statistic decreases. It should be also noted that the difference in power across the horizons increases as the regressor persistence decreases. Figures A23 and A24 present the corresponding power plots for $\delta = -0.5$ and $\delta = 0$, respectively, yielding very similar conclusions.

–Figures A22–A24 here–

Figures A25 to A27 illustrate the corresponding power plots for $n = 500$, which exhibit very similar patterns to the ones derived from the power plots for $n = 1000$; the power of the long–horizon Wald statistic increases as the true value of A increases, while in each case, the power of the statistic decreases as the predictive horizon increases. The loss of power due to the increase of the horizon becomes relatively bigger when the persistence of the regressor decreases. To the contrary, for a highly persistent regressor, this loss is small, implying that the Wald statistic is very powerful even when very long predictive horizons are considered.

–Figures A25–A27 here–

3. Additional Empirical Results

3.1 Definitions of variables

Table A4 contains the definitions of the variables used as predictors in this study and an indicative list of prior studies that have examined their predictive ability. In particular, we consider the following twelve variables: T–bill rate **(tbl)**, long–term yield **(lty)**, term spread **(tms)**, default yield spread **(dfy)**, dividend–price ratio **(d/p)**, dividend yield **(d/y)**, earnings–price ratio **(e/p)**, dividend payout ratio **(d/e)**, book–to–market value ratio **(b/m)**, net equity expansion **(ntis)**, inflation rate **(inf)** and consumption–wealth ratio **(cay)**.

–Table A4 here–

3.2 Regressors' persistence properties with annual data

Table A5 reports the least squares point estimate of the autoregressive root $\widehat{R}_n$ and the results of four unit root tests (ADF, DF–GLS, PP and KPSS) for each of the 12 predictors considered in this study. The results are very similar to the ones reported for monthly and quarterly data in the main body of the study. The autoregressive roots of d/p, d/y, e/p, lty and b/m are remarkably close to unity, while tbl, d/e and dfy also exhibit autoregressive roots greater than 0.9, demonstrating the very high degree of persistence of these regressors at the annual frequency too. On the other hand, tms, cay and ntis, appear to be somewhat less persistent relative to the quarterly frequency. Most importantly, the ambiguity regarding the predictors' order of integration remains unresolved and the unit root tests lead to conflicting conclusions for most of them. There seems to be agreement only on that lty and d/y have a unit root and that dfy and cay are stationary. This evidence further highlights the difficulty in modeling the exact type of persistence of these predictors at any frequency. The proposed IVX–Wald test sidesteps this problem by yielding robust inference with respect to their (uncertain) time series properties.

–Table A5 here–

3.3 Predictability tests with annual data

3.3.1 1–period results This subsection presents the results from univariate predictability tests using annual data for each of the 12 employed predictors. Results are reported in Table A6 for the full sample period, 1927–2012, with the exception of cay which becomes available after 1945. Our findings are very similar to the ones reported for quarterly data in the main body of the study. In particular, according to the IVX–Wald test, we find that d/y, e/p and b/m are significant predictors at the 10% level, while ntis is significant at the 5% level. Moreover, we confirm that cay is highly significant using annual data too. Comparing these results with the inference derived from CY's Q–test, the most important difference is that the latter would fail to indicate the significance of e/p and b/m. With respect to the standard t–ratio, the main difference is that it would additionally indicate d/p as significant at the 10% level and b/m at the 5% level due to its tendency to overreject the null of no predictability, especially for predictors exhibiting a high degree of endogeneity. Finally, our main difference with the inference derived from the JM statistic refers to the level of significance for various predictors. Most importantly, the JM test would find cay to be significant only at the 10% level (p–value: 0.08); it would additionally indicate dfy to be marginally significant, while we find this variable to be insignificant.

–Table A6 here–

3.3.2 Long–horizon results This subsection presents the results from long–horizon univariate predictability tests using annual data. We use predictive horizons up to 5 years. Results are reported in Table A7. Panel A contains the full sample period results, while Panel B reports the corresponding results for the post–1952 period. With respect to the full sample period, we find that predictability becomes weaker, not stronger, as the predictive horizon increases. The only exception is tms which becomes marginally significant as we examine horizons beyond two years. None of the examined predictors is significant at the 5% level when we consider horizons beyond three years. Moreover, in comparison to the 1–year results in Table A6, d/y, e/p and b/m, which were significant at the 10% level, eventually become insignificant for longer predictive horizons. The results for the post–1952 period are even more striking. We find no evidence of predictability at any horizon beyond one year with two exceptions: *i*) cay remains significant at the 5% level even as we increase the horizon to 5 years and *ii*) d/e is significant at longer horizons, but only marginally. Overall, results based on annual data corroborate the results in the main body of the study. Predictability becomes overall weaker, not stronger, as the examined horizon increases, while it almost disappears in the post–1952 period, with the exception of cay, which is the only predictor that remains significant at the 5% level for all horizons considered.

–Table A7 here–

References

[1] Ang, A., Bekaert, G., 2007. Stock return predictability: Is it there?. Review of Financial Studies 20, 651-707.

[2] Avramov, D., 2002. Stock Return Predictability and Model Uncertainty. Journal of Financial Economics 64, 423–458.

[3] Billingsley, P., 1968. Convergence of probability measures. Wiley.

[4] Boudoukh, J., Michaely, R., Richardson, M.P., Roberts, M.R., 2007. On the importance of measuring payout yield: implications for empirical asset pricing. Journal of Finance 62, 877–915.

[5] Campbell, J.Y., 1987. Stock returns and the Term Structure. Journal of Financial Economics 18, 373–399.

[6] Campbell, J.Y, Shiller, R.J., 1988. Stock prices, earnings, and expected dividends. Journal of Finance 43, 661–676.

[7] Campbell J.Y., Thompson, S.B., 2008. Predicting the Equity Premium Out of Sample: Can anything Beat the Historical Average?. Review of Financial Studies 21, 1509-1531.

[8] Campbell J.Y.,Yogo, M., 2006. Efficient tests of stock return predictability. Journal of Financial Economics 81, 27-60.

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

[10] Fama, E.F., French, K.R., 1988. Dividend yields and expected stock returns. Journal of Financial Economics 22, 3–24.

[11] Fama, E.F., French, K.R., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25, 23–49.

[12] Fama, E.F., Schwert, G.W., 1977. Asset returns and inflation. Journal of Financial Economics 5, 115-146.

[13] Goyal, A., Welch, I., 2003. Predicting the equity premium with dividend ratios. Management Science 49, 639-654.

[14] Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Application. Academic Press.

[15] Hodrick, R.J., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. Review of Financial Studies 5, 257–86.

[16] Kallenberg O., 2002. Foundations of Modern Probability. Springer.

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

[18] Kothari, S., Shanken, J., 1997. Book-to-market, dividend yield, and expected market returns: a time-series analysis. Journal of Financial Economics 44, 169–203.

[19] Lamont, O., 1998. Earnings and expected returns. Journal of Finance 53, 1563–1587.

[20] Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. Journal of Finance 56, 815-849.

[21] Lettau, M., Ludvigson, S., 2005. Expected returns and expected dividend growth. Journal of Financial Economics 76, 583–626.

[22] Lewellen, J., 2004. Predicting returns with financial ratios. Journal of Financial Economics 74, 209–235.

[23] Magdalinos, T., 2014. Conditional heteroskedasticity in stochastic regression. Working paper, University of Southampton.

[24] Magdalinos, T., Phillips, P.C.B., 2009. Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. Econometric Theory 25, 482-526.

[25] Pesaran, H.M., Timmermann, A., 1995. Predictability of stock returns: robustness and economic significance. Journal of Finance 50, 1201–1228.

[26] Phillips, P.C.B., Magdalinos T., 2009. Econometric inference in the vicinity of unity. CoFie Working Paper No 7, Singapore Management University.

[27] Pontiff, J., Schall, L.D., 1998. Book-to-market ratios as predictors of market returns. Journal of Financial Economics 49, 141–60.

[28] Phillips, P. C. B., Solo, V., 1992. Asymptotics for Linear Processes. Annals of Statistics 20, 971–1001.

[29] Rozeff, M.S., 1984. Dividend yields are equity risk premiums. Journal of Portfolio Management 11, 68–75.

[30] Stambaugh, R.F., 1999. Predictive regressions. Journal of Financial Economics 54, 375-421.

[31] Torous, W., Valkanov, R., Yan, S., 2004. On predicting stock returns with nearly integrated explanatory variables. Journal of Business 77, 937–966.

[32] Welch, I., Goyal, A., 2008. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. Review of Financial Studies 21, 1455-1508.

[33] Wolf, M., 2000. Stock returns and dividend yields revisited: A new way to look at an old problem. Journal of Business and Economic Statistics 18, 18-30.

Table A1

Finite−sample sizes with autocorrelation coefficient $\phi = 0.25$ in the residuals of the autoregression

This table presents finite−sample sizes, testing the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) when the autocorrelation coefficient in the residuals of the autoregression (23) is $\phi = 0.25$. $W_{0.05}$ corresponds to the rejection rate for the Wald statistic, defined in (19), with 5% nominal size and $Q_{0.05}$ corresponds to the rejection rate resulting from the 95% confidence interval for the Campbell and Yogo (2006) *Q*-test. Results are reported for different degrees of correlation between the residuals of regressions (22) and (23) in the main body of the study, $\delta = -0.95$, −0.5, 0, 0.5 and 0.95, different sample sizes *n*= 100, 250, 500 and 1,000 and for different local-to-unity parameters *C*= 0, −5, −10, −15, −20 and −50, which in each sample size case correspond to different autoregressive roots (R_n) reported in the third column. The reported results are based on the Monte Carlo simulation described in Section 2.1 of the main body of the study and the average rejection rates are calculated over 10,000 repetitions.

			$\delta = -0.95$		$\delta = -0.50$		$\delta = 0$		$\delta = 0.50$		$\delta = 0.95$	
n	*C*	R_n	$W_{0.05}$	$Q_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$W_{0.05}$	$Q_{0.05}$
100	0	1.000	0.062	0.048	0.062	0.040	0.045	0.047	0.062	0.039	0.066	0.050
	−5	0.950	0.067	0.045	0.066	0.039	0.054	0.049	0.061	0.035	0.073	0.053
	−10	0.900	0.064	0.051	0.060	0.036	0.056	0.049	0.061	0.036	0.065	0.053
	−20	0.800	0.055	0.071	0.059	0.040	0.058	0.051	0.058	0.039	0.060	0.073
	−50	0.500	0.057	0.173	0.054	0.065	0.060	0.055	0.059	0.068	0.056	0.170
250	0	1.000	0.061	0.042	0.057	0.036	0.052	0.049	0.054	0.036	0.060	0.043
	−5	0.980	0.063	0.044	0.060	0.033	0.053	0.049	0.056	0.031	0.066	0.045
	−10	0.960	0.062	0.044	0.056	0.032	0.051	0.051	0.054	0.029	0.062	0.047
	−20	0.920	0.064	0.071	0.052	0.028	0.054	0.049	0.054	0.031	0.063	0.061
	−50	0.800	0.055	0.218	0.052	0.063	0.049	0.047	0.055	0.063	0.053	0.210
500	0	1.000	0.072	0.054	0.066	0.044	0.050	0.051	0.061	0.039	0.073	0.054
	−5	0.990	0.072	0.053	0.063	0.040	0.053	0.049	0.062	0.037	0.073	0.053
	−10	0.980	0.068	0.047	0.060	0.036	0.054	0.050	0.061	0.034	0.071	0.052
	−20	0.960	0.063	0.059	0.055	0.032	0.056	0.051	0.056	0.033	0.061	0.056
	−50	0.900	0.053	0.150	0.051	0.053	0.055	0.052	0.056	0.055	0.055	0.155
1000	0	1.000	0.054	0.039	0.056	0.037	0.048	0.048	0.055	0.034	0.055	0.041
	−5	0.995	0.064	0.044	0.057	0.032	0.050	0.050	0.050	0.030	0.060	0.044
	−10	0.990	0.061	0.044	0.056	0.032	0.050	0.048	0.054	0.030	0.062	0.048
	−20	0.980	0.056	0.046	0.055	0.030	0.056	0.053	0.051	0.030	0.060	0.047
	−50	0.950	0.058	0.097	0.055	0.036	0.052	0.049	0.054	0.035	0.052	0.101

Table A2

Finite−sample sizes with autocorrelation coefficient $\phi = -0.1$ in the residuals of the autoregression

This table presents finite−sample sizes, testing the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) when the autocorrelation coefficient in the residuals of the autoregression (23) is $\phi = -0.1$. $W_{0.05}$ corresponds to the rejection rate for the Wald statistic, defined in (19), with 5% nominal size and $Q_{0.05}$ corresponds to the rejection rate resulting from the 95% confidence interval for the Campbell and Yogo (2006) *Q*-test. Results are reported for different degrees of correlation between the residuals of regressions (22) and (23) in the main body of the study, $\delta = -0.95$, −0.5, 0, 0.5 and 0.95, different sample sizes *n*= 100, 250, 500 and 1,000 and for different local-to-unity parameters *C*= 0, −5, −10, −15, −20 and −50, which in each sample size case correspond to different autoregressive roots (R_n) reported in the third column. The reported results are based on the Monte Carlo simulation described in Section 2.1 of the main body of the study and the average rejection rates are calculated over 10,000 repetitions.

			$\delta = -0.95$		$\delta = -0.50$		$\delta = 0$		$\delta = 0.50$		$\delta = 0.95$	
n	C	R_n	$W_{0.05}$	$Q_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$W_{0.05}$	$Q_{0.05}$
100	0	1.000	0.067	0.053	0.063	0.041	0.052	0.052	0.062	0.042	0.065	0.048
	−5	0.950	0.067	0.056	0.058	0.038	0.053	0.047	0.060	0.039	0.072	0.063
	−10	0.900	0.064	0.064	0.063	0.044	0.056	0.050	0.057	0.040	0.064	0.064
	−20	0.800	0.060	0.098	0.052	0.041	0.054	0.047	0.057	0.046	0.061	0.103
	−50	0.500	0.055	0.323	0.056	0.126	0.054	0.049	0.053	0.119	0.055	0.306
250	0	1.000	0.063	0.044	0.060	0.037	0.051	0.051	0.053	0.034	0.055	0.040
	−5	0.980	0.061	0.045	0.058	0.036	0.054	0.050	0.055	0.033	0.063	0.045
	−10	0.960	0.062	0.051	0.052	0.029	0.054	0.051	0.056	0.036	0.058	0.047
	−20	0.920	0.057	0.071	0.053	0.034	0.056	0.051	0.056	0.035	0.058	0.070
	−50	0.800	0.052	0.217	0.054	0.067	0.051	0.048	0.047	0.062	0.052	0.225
500	0	1.000	0.059	0.042	0.056	0.037	0.050	0.049	0.052	0.036	0.058	0.041
	−5	0.990	0.062	0.046	0.056	0.032	0.052	0.053	0.050	0.030	0.061	0.044
	−10	0.980	0.065	0.046	0.055	0.032	0.050	0.049	0.052	0.031	0.062	0.045
	−20	0.960	0.060	0.053	0.050	0.030	0.054	0.053	0.051	0.029	0.057	0.055
	−50	0.900	0.052	0.164	0.048	0.047	0.050	0.048	0.054	0.053	0.056	0.171
1000	0	1.000	0.059	0.042	0.053	0.031	0.050	0.049	0.052	0.037	0.054	0.040
	−5	0.995	0.063	0.043	0.055	0.030	0.050	0.049	0.054	0.030	0.062	0.046
	−10	0.990	0.058	0.044	0.053	0.030	0.051	0.050	0.054	0.030	0.052	0.041
	−20	0.980	0.058	0.045	0.053	0.030	0.052	0.051	0.057	0.032	0.057	0.049
	−50	0.950	0.051	0.127	0.051	0.040	0.051	0.048	0.054	0.040	0.052	0.126

Table A3

Finite−sample sizes using a conditionally heteroskedastic DGP for stock returns

This table presents finite−sample sizes, testing the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (34) of the Online Appendix when the DGP for stock returns is conditionally heteroskedastic and there is no autocorrelation ($\phi = 0$) in the residuals of the autoregression. In particular, the residuals of the predictive regression for stock returns are conditionally heteroskedastic, following a GARCH (1,1) process. The employed parameter values are derived from fitting a GARCH (1,1) to the residuals estimated from regressing S&P 500 value-weighted log excess returns on dividend yield, using monthly data for the period 1927−2012. $W_{0.05}$ corresponds to the rejection rate for the Wald statistic, defined in (19), with 5% nominal size and $Q_{0.05}$ corresponds to the rejection rate resulting from the 95% confidence interval for the Campbell and Yogo (2006) Q-test. Results are reported for different degrees of correlation between the residuals, $\delta = -0.95$, −0.5, 0, 0.5 and 0.95, different sample sizes n= 100, 250, 500 and 1,000 and for different local-to-unity parameters C= 0, −5, −10, −15, −20 and −50, which in each sample size case correspond to different autoregressive roots (R_n) reported in the third column. The reported results are based on the Monte Carlo simulation described in Section 2.7 of the Online Appendix and the average rejection rates are calculated over 10,000 repetitions.

			$\delta = -0.95$		$\delta = -0.50$		$\delta = 0$		$\delta = 0.50$		$\delta = 0.95$	
n	C	R_n	$W_{0.05}$	$Q_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$W_{0.05}$	$Q_{0.05}$	$W_{0.05}$	$Q_{0.05}$
100	0	1.000	0.066	0.046	0.062	0.041	0.051	0.053	0.064	0.043	0.062	0.040
	−5	0.950	0.058	0.044	0.058	0.039	0.054	0.049	0.059	0.041	0.061	0.037
	−10	0.900	0.057	0.041	0.059	0.039	0.054	0.048	0.060	0.041	0.058	0.039
	−20	0.800	0.056	0.050	0.055	0.043	0.055	0.051	0.058	0.048	0.053	0.050
	−50	0.500	0.050	0.140	0.051	0.093	0.052	0.051	0.050	0.091	0.050	0.135
250	0	1.000	0.053	0.036	0.053	0.038	0.051	0.052	0.059	0.043	0.057	0.036
	−5	0.980	0.058	0.034	0.051	0.032	0.053	0.051	0.053	0.035	0.056	0.033
	−10	0.960	0.052	0.030	0.051	0.033	0.053	0.050	0.053	0.035	0.058	0.033
	−20	0.920	0.047	0.031	0.049	0.030	0.052	0.048	0.056	0.036	0.047	0.032
	−50	0.800	0.048	0.070	0.045	0.056	0.056	0.055	0.047	0.053	0.045	0.067
500	0	1.000	0.053	0.032	0.048	0.034	0.050	0.052	0.055	0.035	0.049	0.033
	−5	0.990	0.055	0.032	0.051	0.032	0.048	0.047	0.054	0.034	0.053	0.031
	−10	0.980	0.049	0.028	0.053	0.034	0.051	0.050	0.051	0.032	0.053	0.030
	−20	0.960	0.050	0.030	0.049	0.030	0.053	0.050	0.053	0.033	0.047	0.028
	−50	0.900	0.045	0.043	0.048	0.041	0.049	0.048	0.047	0.040	0.044	0.042
1000	0	1.000	0.048	0.033	0.051	0.035	0.048	0.048	0.052	0.036	0.047	0.031
	−5	0.995	0.056	0.029	0.051	0.032	0.055	0.054	0.058	0.036	0.054	0.029
	−10	0.990	0.054	0.030	0.056	0.034	0.049	0.049	0.052	0.031	0.049	0.027
	−20	0.980	0.047	0.024	0.046	0.027	0.057	0.054	0.050	0.033	0.045	0.026
	−50	0.950	0.043	0.028	0.046	0.031	0.053	0.053	0.043	0.028	0.044	0.029

Table A4
Definitions of predictive regressors
This table reports the variables used as predictive regressors in this study, their definition and some prior studies that have examined their predictive ability.

Variable	Definition	Indicative list of prior studies
Dividend payout ratio (d/e)	Difference between the log of dividends and the log of earnings	Lamont (1998)
Earnings-price ratio (e/p)	Difference between the log of earnings and the log of stock prices. Earnings are calculated using a 12-month rolling sum of earnings of S&P 500 companies	Campbell and Shiller (1988), Fama and French (1988), Pesaran and Timmermann (1995), Lamont (1998), Lewellen (2004), Ang and Bekaert (2007), Campbell and Thompson (2008)
Long-term yield (lty)	Long-term US government bond yield from Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook	Keim and Stambaugh (1986), Fama and French (1989), Pontiff and Schall (1998), Torous et al. (2004), Campbell and Yogo (2006)
T-bill rate (tbl)	3-month US Treasury bill rate taken from FRED. For the period before 1934 it is extracted from the NBER Macrohistory database	Pesaran and Timmermann (1995), Pontiff and Schall (1998), Torous et al. (2004), Campbell and Yogo (2006), Ang and Bekaert (2007), Avramov (2002), Campbell and Thompson (2008)
Term spread (tms)	Difference between the long-term yield and the T-bill rate	As for the long-term yield
Dividend-price ratio (d/p)	Difference between the log of dividends and the log of stock prices. Dividends are calculated using a 12-month rolling sum of dividends paid on the S&P 500 index	Rozeff (1984), Campbell (1987), Campbell and Shiller (1988), Fama and French (1988), Hodrick (1992), Lamont (1998), Stambaugh (1999), Wolf (2000), Goyal and Welch (2003), Lewellen (2004), Torous et al. (2004), Lettau and Ludvigson (2005), Campbell and Yogo (2006), Ang and Bekaert (2007), Campbell and Thompson (2008)
Dividend yield (d/y)	Difference between the log of dividends and the log of lagged stock prices	
Default yield spread (dfy)	Difference between the BAA and AAA-rated corporate bond yields taken from FRED	Fama and French (1989), Avramov (2002), Torous et al. (2004), Campbell and Thompson (2008)
Book-to-market value ratio (b/m)	Ratio of book value to market value for the DJIA	Kothari and Shanken (1997), Pontiff and Schall (1998), Avramov (2002), Lewellen (2004), Campbell and Thompson (2008)
Net equity expansion (ntis)	Ratio of the 12-month moving sum of net equity issues by NYSE listed stocks divided by the total end-of-year market capitalization of these stocks	Boudoukh et al. (2007) use net payout yield, Welch and Goyal (2008)
Inflation rate (inf)	Based on the Consumer Price Index from the Bureau of Labor Statistics	Fama and Schwert (1977), Fama (1981), Welch and Goyal (2008)
Consumption-wealth ratio (cay)	Transitory deviation of consumption from its cointegrating relationship with asset holdings and labor income	Lettau and Ludvigson (2001), Welch and Goyal (2008)

Table A5
Unit root tests for predictive regressors−Annual data

This table presents the results of unit root tests for the following list of financial and economic variables defined in Section 3 of the main body of the study: Dividend payout ratio (d/e), long-term yield (lty), dividend yield (d/y), dividend price ratio (d/p), T-bill rate (tbl), earnings price ratio (e/p), book-to-market value ratio (b/m), default yield spread (dfy), net equity expansion (ntis), term spread (tms), inflation rate (inf) and consumption-wealth ratio (cay). $\hat{R}_n$ corresponds to the least squares point estimate of the AR(1): $x_t = R_n x_{t-1} + u_t$. ADF stands for the augmented Dickey-Fuller test statistic, DF-GLS refers to the Elliot et al. (1996) Dickey-Fuller-GLS test statistic, PP stands for the Phillips-Perron test statistic and KPSS refers to the Kwiatkowski et al. (1992) test statistic. The Bayesian Information Criterion has been been used to select the optimal lag length for ADF and DF-GLS test statistics. The sample period is 1927−2012. *, ** and *** imply rejection of the null hypothesis of a unit root (for ADF, DF-GLS and PP) or stationarity (for KPSS) at 10%, 5% and 1% level respectively.

	$\hat{R}_n$	ADF	DF-GLS	PP	KPSS
Dividend payout ratio	0.923	−4.498***	−4.525***	−4.403***	1.026***
Long-term yield	0.989	−1.259	−1.227	−1.153	0.528**
Dividend yield	1.000	−1.621	−1.223	−1.667	0.764***
Dividend-price ratio	1.000	−2.219	−1.884*	−2.025	0.901***
T-bill rate	0.953	−1.763	−1.809*	−1.894	0.456*
Earnings-price ratio	0.996	−3.617***	−3.180***	−3.570***	0.407*
Book-to-market value ratio	0.973	−2.606*	−2.484**	−2.614*	0.515**
Default yield spread	0.913	−3.758***	−3.719***	−3.787***	0.223
Net equity expansion	0.681	−5.128***	−2.877***	−4.891***	0.633**
Term spread	0.829	−4.565***	−4.064***	−4.234***	0.350*
Inflation rate	0.771	−2.230	−1.477	−4.213***	0.325
Consumption-wealth ratio	0.706	−3.521**	−2.142**	−3.827***	0.120

Table A6
Univariate predictive regressions−Annual data

This table presents the results of univariate predictive regression models, as in equation (1), during the sample period 1927−2012. The dependent variable is the annual S&P 500 value-weighted log excess returns and the lagged persistent regressor is each of the following variables defined in Section 3 of the main body of the study: Dividend payout ratio (d/e), long-term yield (lty), dividend yield (d/y), dividend price ratio (d/p), T-bill rate (tbl), earnings price ratio (e/p), book-to-market value ratio (b/m), default yield spread (dfy), net equity expansion (ntis), term spread (tms), inflation rate (inf) and consumption-wealth ratio (cay). The sample period for cay is 1945−2012. $\tilde{A}_{OLS}$ stands for the least squares slope coefficient estimated via regression model (1), while t_{OLS} is the corresponding t-statistic under the null hypothesis that A is equal to zero (i.e., no predictability). $\tilde{A}_{IVX}$, defined in (17), stands for the slope coefficient for the predictive regression (16) estimated via the proposed instrumental variable (IVX) approach, while IVX-Wald refers to the Wald statistic, defined in equation (19), under the null hypothesis that the slope coefficient A is equal to zero. δ denotes the correlation coefficient between the residuals of regression models (1) and (2). *, ** and *** imply rejection of the null hypothesis at 10%, 5% and 1% level respectively. CY 90% CI stands for the 90% Bonferroni confidence interval for the bias-corrected scaled least squares slope coefficient of the predictive regression using the *Q*-test of Campbell and Yogo (2006). Bold fonts indicate rejection of the null hypothesis of no predictability at the 10% level. JM reports the p-value for the $\pi^{*}_{0.05}$ statistic of Jansson and Moreira (2006) under the null hypothesis of no predictability.

Regressors	$\tilde{A}_{OLS}$	t_{OLS}	$\tilde{A}_{IVX}$	IVX-Wald	δ	CY 90% CI		JM
Dividend payout ratio	0.0142	0.21	0.0059	0.008	−0.325	−0.146	0.188	0.15
Long-term yield	−0.4464	−0.56	−0.4494	0.301	−0.044	−0.070	0.038	0.40
Dividend yield	0.0887	1.84*	0.0985	3.579*	0.051	**0.004**	**0.130**	0.02**
Dividend-price ratio	0.0775	1.67*	0.0823	2.635	−0.816	−0.026	0.181	0.42
T-bill rate	−0.7904	−1.13	−0.7593	1.154	0.127	−0.142	0.029	0.35
Earnings-price ratio	0.0878	1.69*	0.0882	2.849*	−0.248	−0.001	0.268	0.03**
Book-to-market value ratio	0.1718	2.11**	0.1571	3.646*	−0.797	−0.000	0.235	0.03**
Default yield spread	0.5537	0.20	0.0789	0.001	−0.626	−0.117	0.204	0.09*
Net equity expansion	−1.6430	−2.09**	−2.0374	4.660**	0.101	**−0.356**	**−0.046**	0.05*
Term spread	2.1038	1.38	2.0318	1.755	−0.135	−0.029	0.271	0.33
Inflation rate	0.1704	0.32	0.1991	0.131	−0.024	−0.121	0.127	0.47
Consumption-wealth ratio	2.3950	2.88***	2.6013	8.238***	−0.408	**0.008**	**0.253**	0.08*

Table A7

Long−horizon univariate predictive regressions−Annual data

This table presents the results of long-horizon univariate predictive regression models, as in equation (30), during the sample periods 1927−2012 (Panel A) and 1952−2012 (Panel B), for various horizons (*K*-yrs). The dependent variable is the cumulative S&P 500 value-weighted log excess return from year *t* to year *t+K-1*, corresponding to a horizon of *K* years, and the lagged persistent regressor is each of the following variables defined in Section 3 of the main body of the study: Dividend payout ratio (d/e), long-term yield (lty), dividend yield (d/y), dividend price ratio (d/p), T-bill rate (tbl), earnings price ratio (e/p), book-to-market value ratio (b/m), default yield spread (dfy), net equity expansion (ntis), term spread (tms), inflation rate (inf) and consumption-wealth ratio (cay). The table reports the long-horizon Wald statistic, defined in equation (34), under the null hypothesis that the slope coefficient of the long-horizon univariate predictive regression estimated via the proposed instrumental variable (IVX) approach, is equal to zero (i.e., no predictability). *, ** and *** imply rejection of the null hypothesis at 10%, 5% and 1% level respectively.

Panel A: 1927−2012												
K-yrs	d/e	lty	d/y	d/p	tbl	e/p	b/m	dfy	ntis	tms	inf	
2	0.628	0.088	2.986*	3.882**	0.851	2.626	3.698*	0.189	6.540**	2.435	0.138	
3	0.503	0.066	2.511	3.292*	0.789	2.540	3.140*	0.117	4.821**	2.802*	0.151	
4	0.735	0.049	2.074	3.320*	0.700	2.469	3.124*	0.262	3.532*	3.135*	0.194	
5	0.640	0.026	1.882	2.997*	0.499	2.431	2.440	0.185	2.841*	2.870*	0.326	
Panel B: 1952−2012												
K-yrs	d/e	lty	d/y	d/p	tbl	e/p	b/m	dfy	ntis	tms	inf	cay
2	3.122*	0.032	1.200	1.855	0.680	0.507	0.256	0.375	0.269	2.224	0.716	7.772***
3	3.151*	0.022	1.239	1.350	0.410	0.469	0.049	0.015	0.092	1.510	0.704	6.113**
4	4.534**	0.000	1.308	1.125	0.289	0.339	0.044	0.093	0.048	1.982	0.692	5.581**
5	3.306*	0.052	1.159	0.986	0.089	0.561	0.100	0.036	0.046	2.246	0.293	4.362**

Figure A1

Power plots for sample size n=1,000 and residuals' correlation coefficient $\delta = -0.95$

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of A increases. The solid curve ($\text{Wald}_{0.05}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line). The dashed curve ($\text{CYQ}_{0.05}$) illustrates the rejection rate using the 95% confidence interval of the Campbell and Yogo (2006) *Q*-test. The dash-dot curve ($\text{JM}_{0.05}$) illustrates the rejection rate using the $\pi^*_{0.05}$ statistic of Jansson and Moreira (2006). Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 with 10,000 repetitions for a sample size of n=1,000, correlation coefficient between the residuals of regressions (22) and (23) $\delta = -0.95$ and no autocorrelation in the residuals of the autoregressive equation, i.e., $\phi = 0$ in (24).

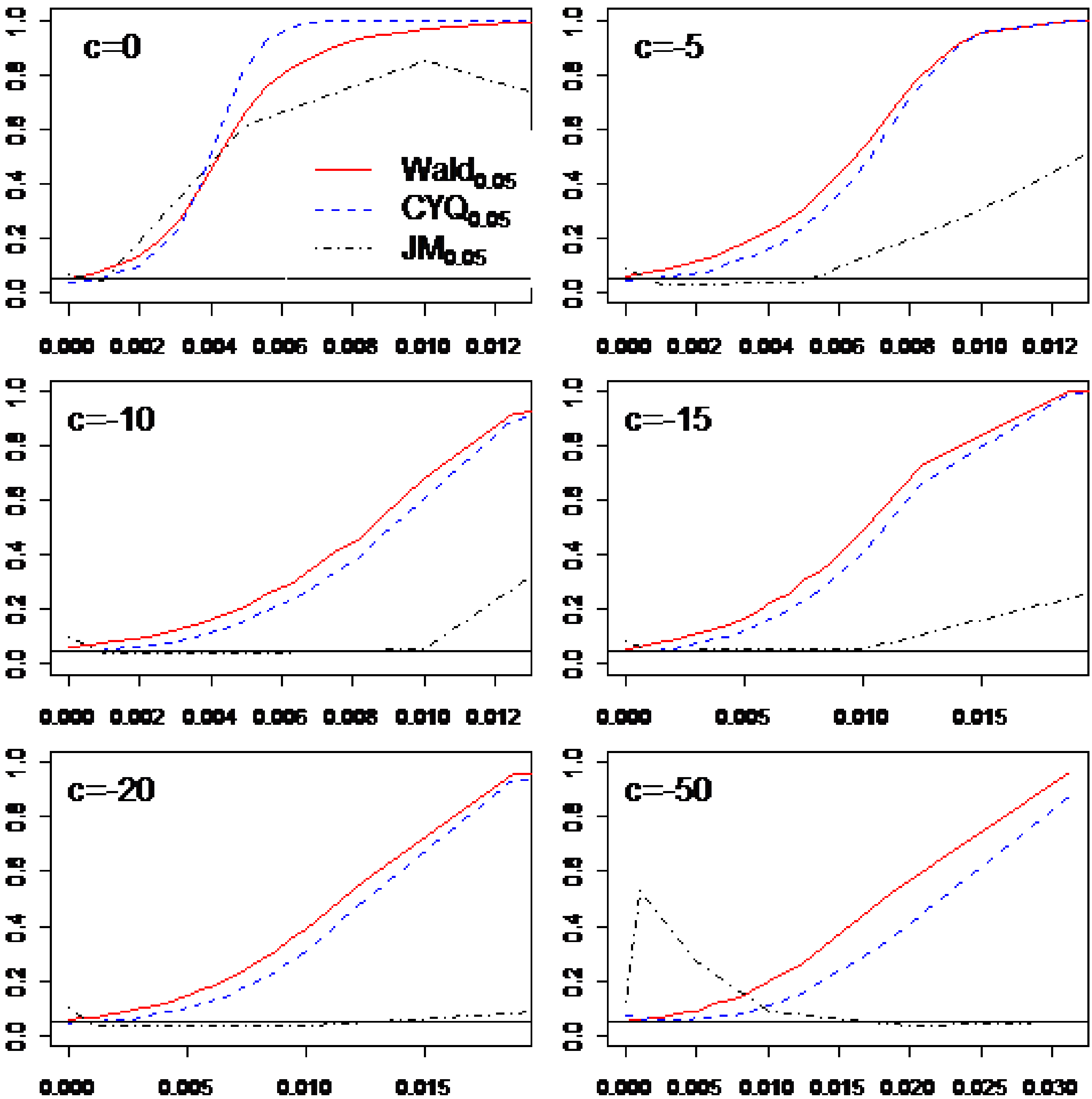


Figure A2

Power plots for sample size n=1,000 and residuals' correlation coefficient $\delta = -0.5$

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of *A* increases. The solid curve ($Wald_{0.05}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line). The dashed curve ($CYQ_{0.05}$) illustrates the rejection rate using the 95% confidence interval of the Campbell and Yogo (2006) *Q*-test. The dash-dot curve ($JM_{0.05}$) illustrates the rejection rate using the $\pi^*_{0.05}$ statistic of Jansson and Moreira (2006). Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 with 10,000 repetitions for a sample size of n=1,000, correlation coefficient between the residuals of regressions (22) and (23) $\delta = -0.5$ and no autocorrelation in the residuals of the autoregressive equation, i.e., $\phi = 0$ in (24).

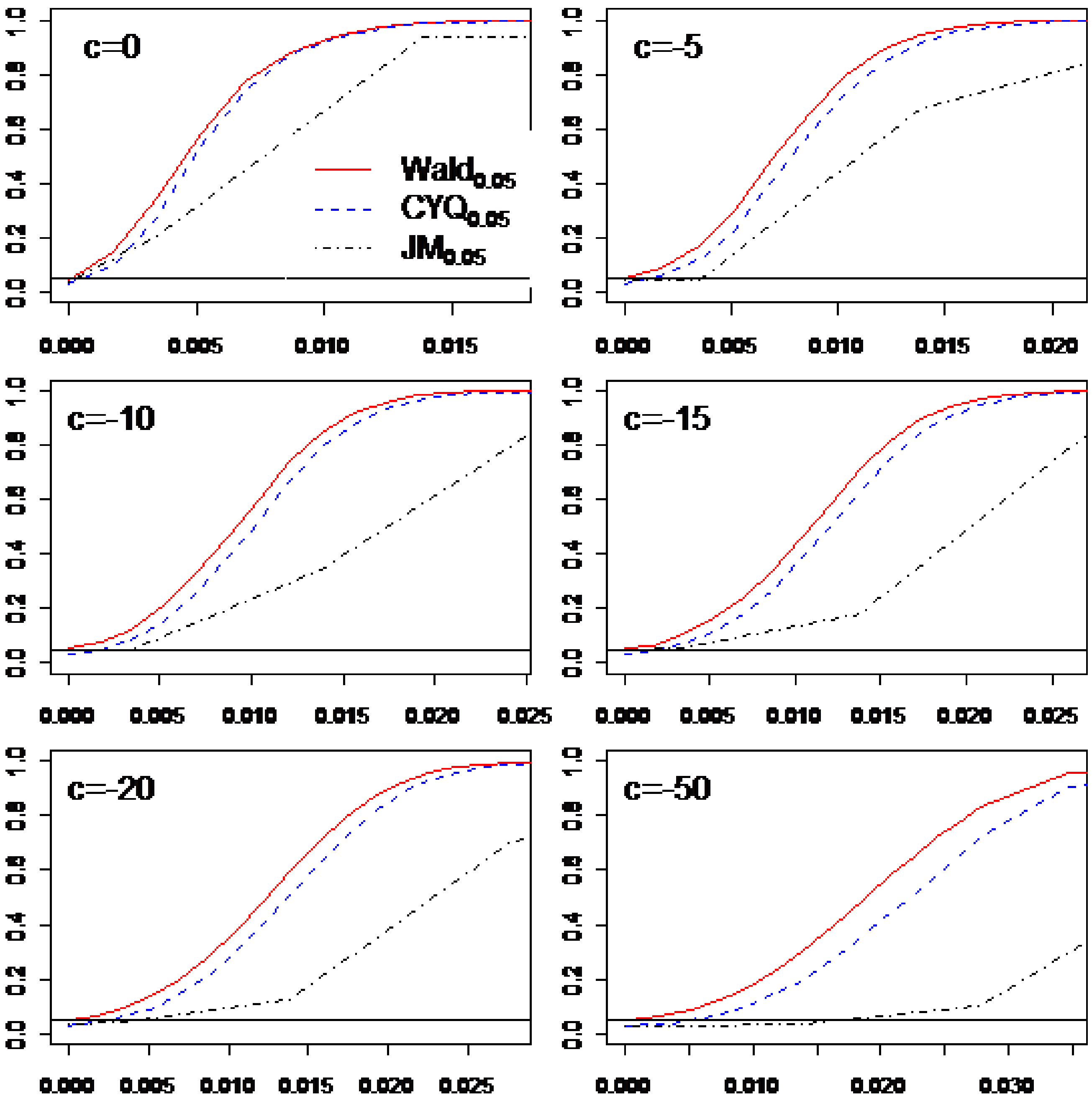


Figure A3

Power plots for sample size n=1,000 and residuals' correlation coefficient $\delta = 0$

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of *A* increases. The solid curve ($\text{Wald}_{0.05}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line). The dashed curve ($\text{CYQ}_{0.05}$) illustrates the rejection rate using the 95% confidence interval of the Campbell and Yogo (2006) *Q*-test. The dash-dot curve ($\text{JM}_{0.05}$) illustrates the rejection rate using the $\pi^*_{0.05}$ statistic of Jansson and Moreira (2006). Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 with 10,000 repetitions for a sample size of n=1,000, correlation coefficient between the residuals of regressions (22) and (23) $\delta = 0$ and no autocorrelation in the residuals of the autoregressive equation, i.e., $\phi = 0$ in (24).

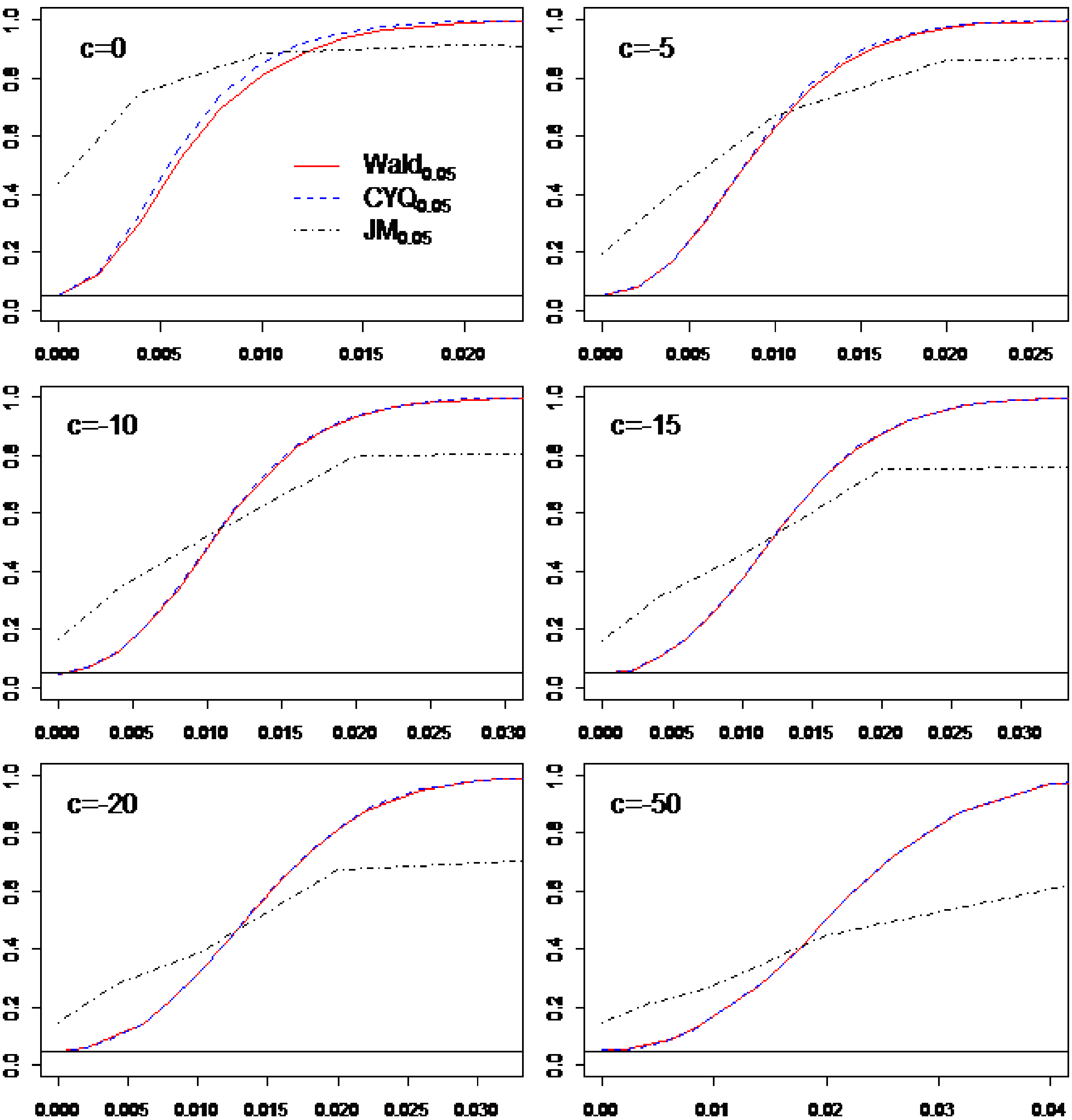


Figure A4

Power plots for sample size n=250, residuals' correlation coefficient $\delta = -0.95$ and autocorrelation coefficient $\phi = 0.5$ in the residuals of the autoregression

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of *A* increases. The solid curve ($\text{Wald}_{0.05}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line). The dashed curve ($\text{CYQ}_{0.05}$) illustrates the rejection rate using the 95% confidence interval of the Campbell and Yogo (2006) *Q*-test. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 of the main body of the study with 10,000 repetitions for a sample size of n=250, correlation coefficient between the residuals of regressions (22) and (23) $\delta = -0.95$ and autocorrelation coefficient $\phi = 0.5$ in the residuals of the autoregression (23).

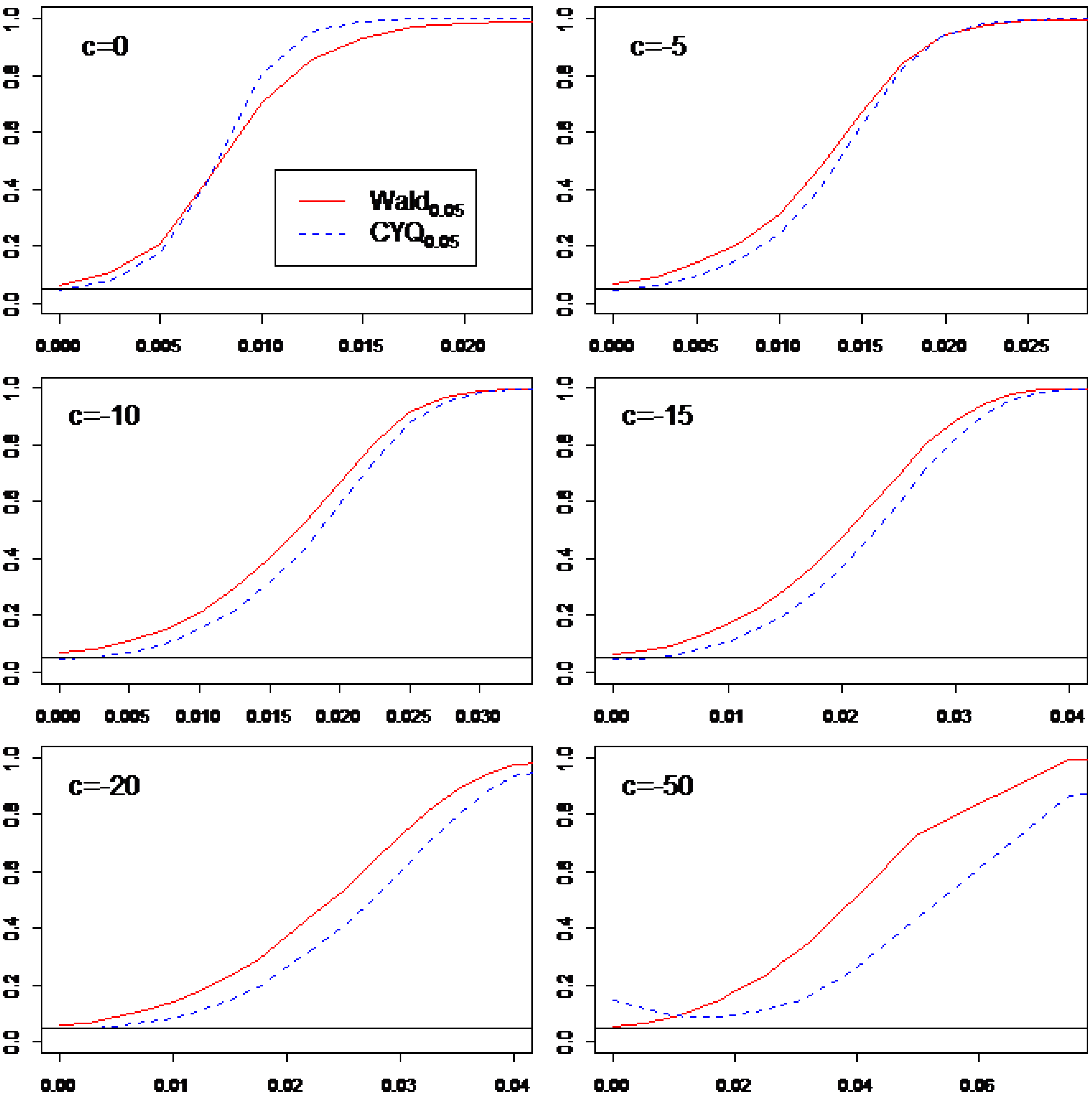


Figure A5

Power plots for sample size n=250, residuals' correlation coefficient $\delta = -0.5$ and autocorrelation coefficient $\phi = 0.5$ in the residuals of the autoregression

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of *A* increases. The solid curve ($Wald_{0.05}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line). The dashed curve ($CYQ_{0.05}$) illustrates the rejection rate using the 95% confidence interval of the Campbell and Yogo (2006) *Q*-test. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 of the main body of the study with 10,000 repetitions for a sample size of n=250, correlation coefficient between the residuals of regressions (22) and (23) $\delta = -0.5$ and autocorrelation coefficient $\phi = 0.5$ in the residuals of the autoregression (23).

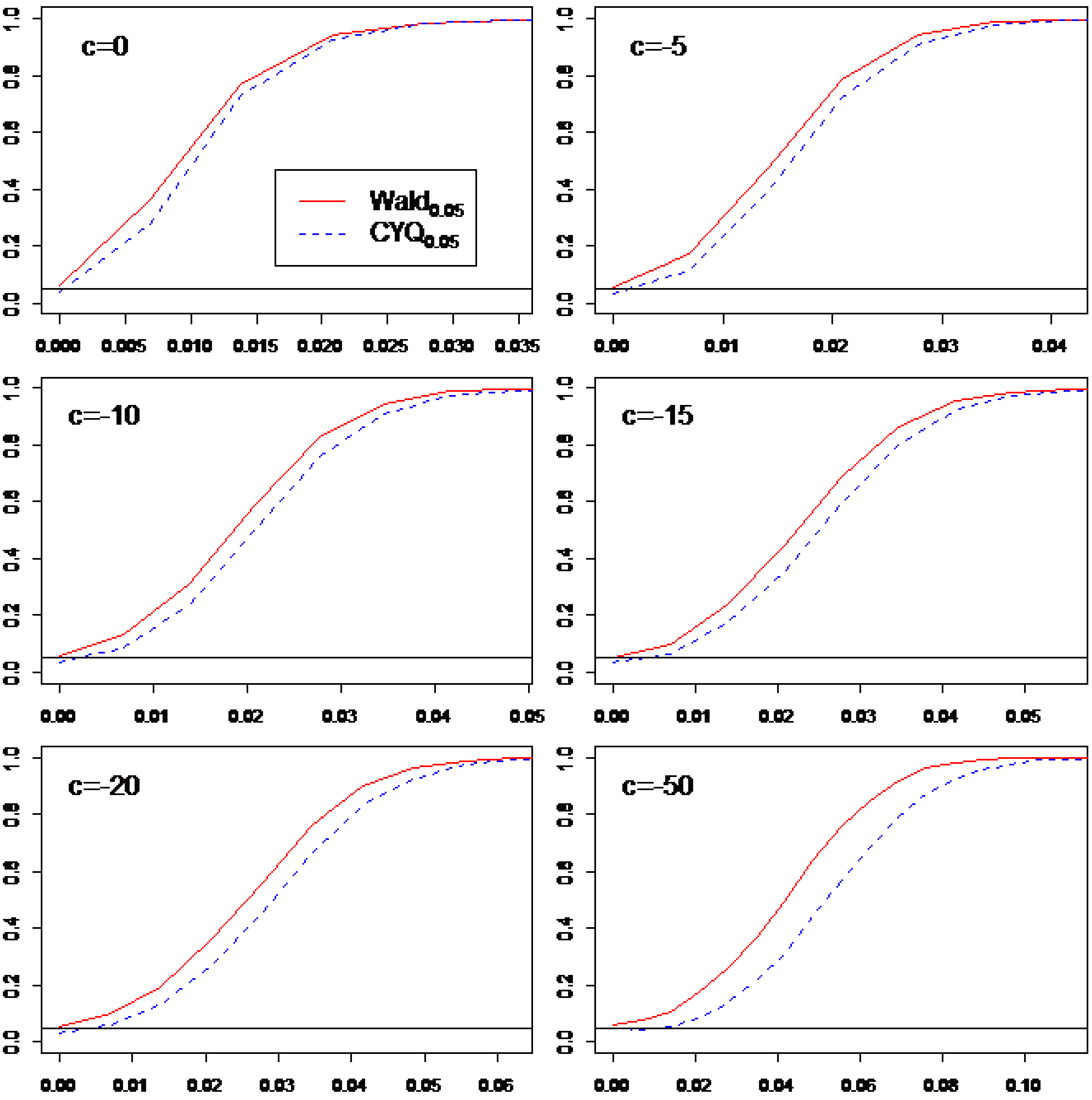


Figure A6

Power plots for sample size n=250, residuals' correlation coefficient $\delta = 0$ and autocorrelation coefficient $\phi = 0.5$ in the residuals of the autoregression

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of *A* increases. The solid curve ($Wald_{0.05}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line). The dashed curve ($CYQ_{0.05}$) illustrates the rejection rate using the 95% confidence interval of the Campbell and Yogo (2006) *Q*-test. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 of the main body of the study with 10,000 repetitions for a sample size of n=250, correlation coefficient between the residuals of regressions (22) and (23) $\delta = 0$ and autocorrelation coefficient $\phi = 0.5$ in the residuals of the autoregression (23).

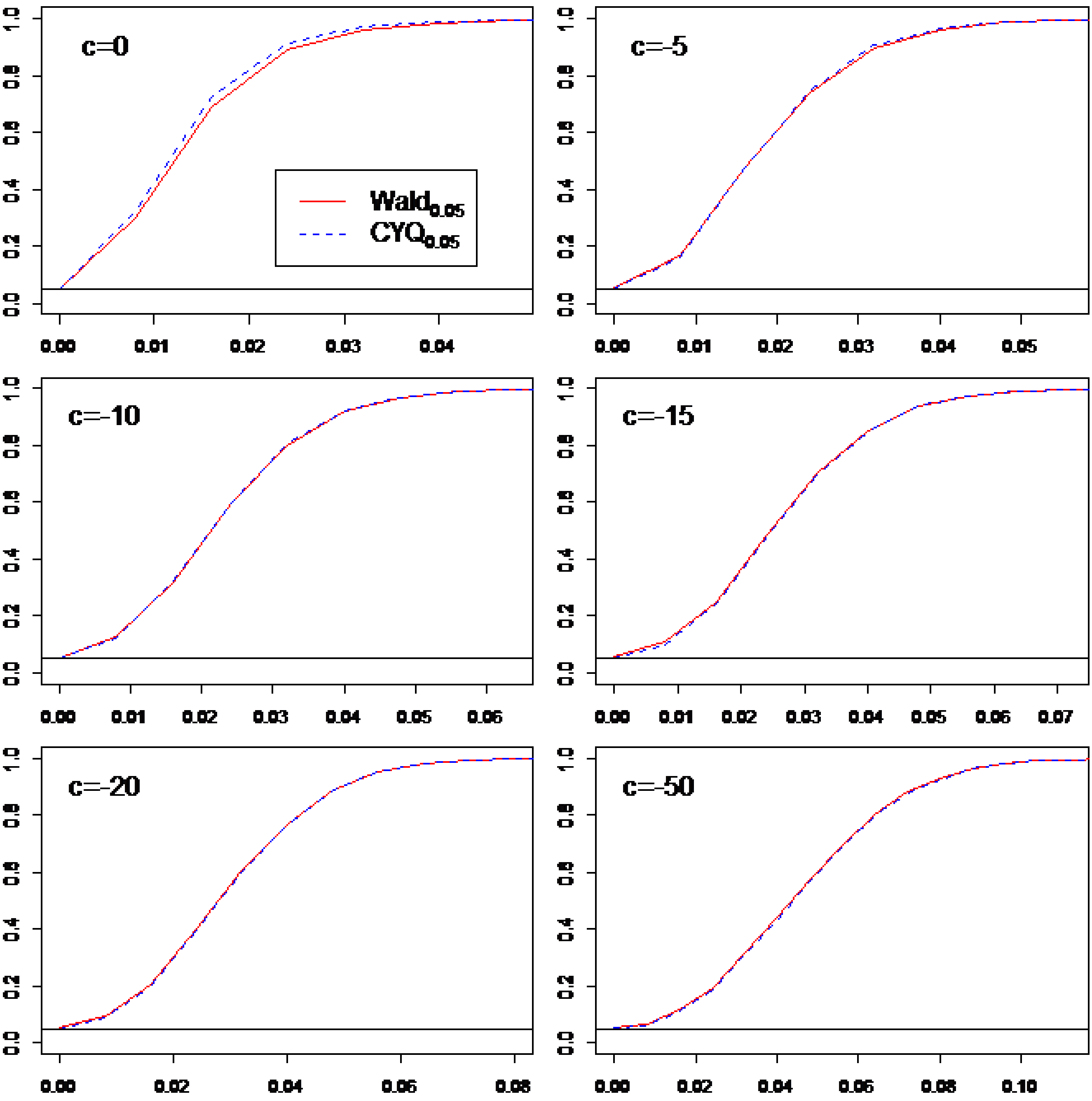


Figure A7

Power plots for sample size n=1,000, residuals' correlation coefficient $\delta = -0.95$ and autocorrelation coefficient $\phi = 0.5$ in the residuals of the autoregression

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of *A* increases. The solid curve ($Wald_{0.05}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line). The dashed curve ($CYQ_{0.05}$) illustrates the rejection rate using the 95% confidence interval of the Campbell and Yogo (2006) *Q*-test. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 of the main body of the study with 10,000 repetitions for a sample size of n=1,000, correlation coefficient between the residuals of regressions (22) and (23) $\delta = -0.95$ and autocorrelation coefficient $\phi = 0.5$ in the residuals of the autoregression (23).

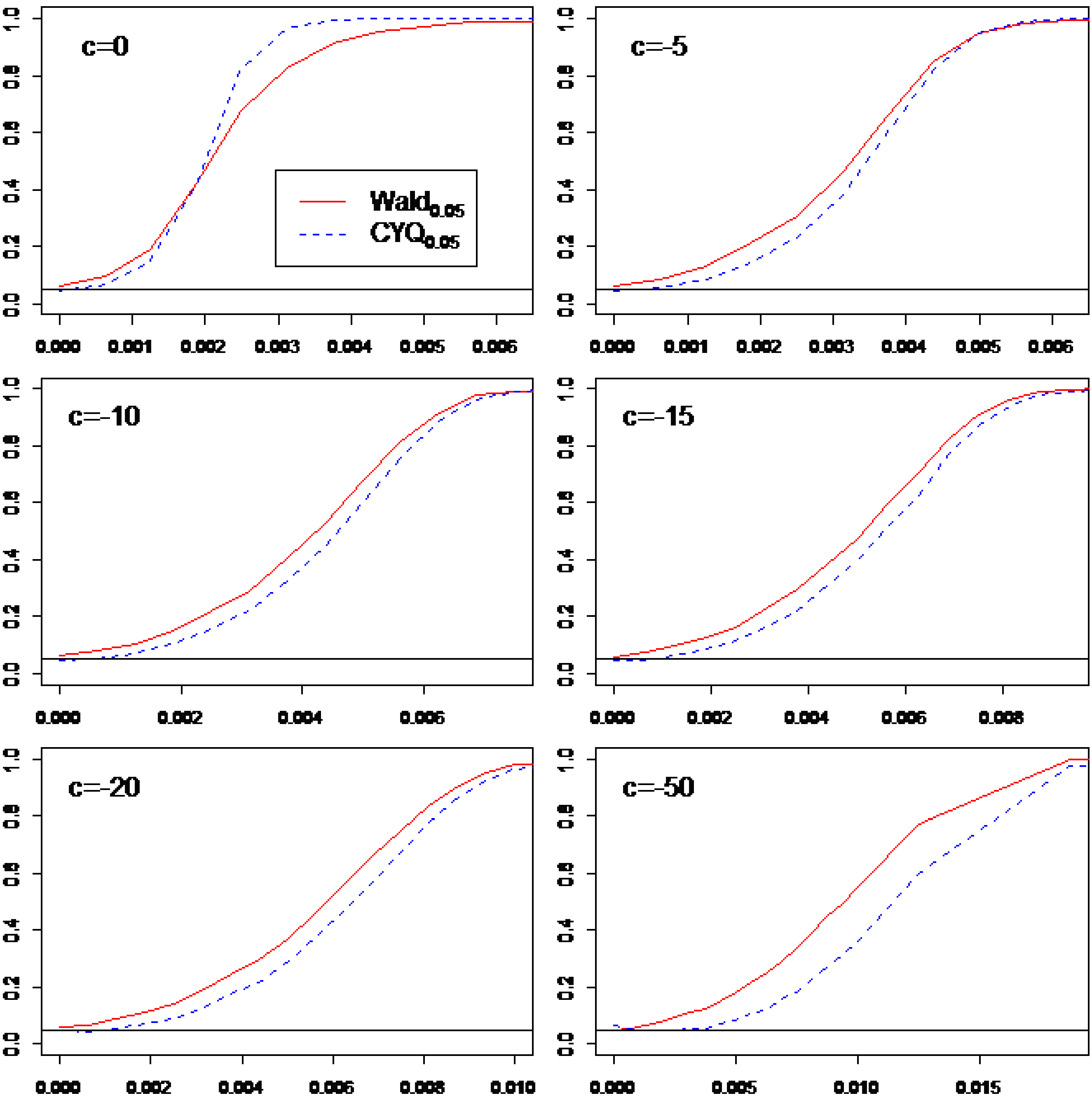


Figure A8

Power plots for sample size n=1,000, residuals' correlation coefficient $\delta = -0.5$ and autocorrelation coefficient $\phi = 0.5$ in the residuals of the autoregression

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of *A* increases. The solid curve ($Wald_{0.05}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line). The dashed curve ($CYQ_{0.05}$) illustrates the rejection rate using the 95% confidence interval of the Campbell and Yogo (2006) *Q*-test. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 of the main body of the study with 10,000 repetitions for a sample size of n=1,000, correlation coefficient between the residuals of regressions (22) and (23) $\delta = -0.5$ and autocorrelation coefficient $\phi = 0.5$ in the residuals of the autoregression (23).

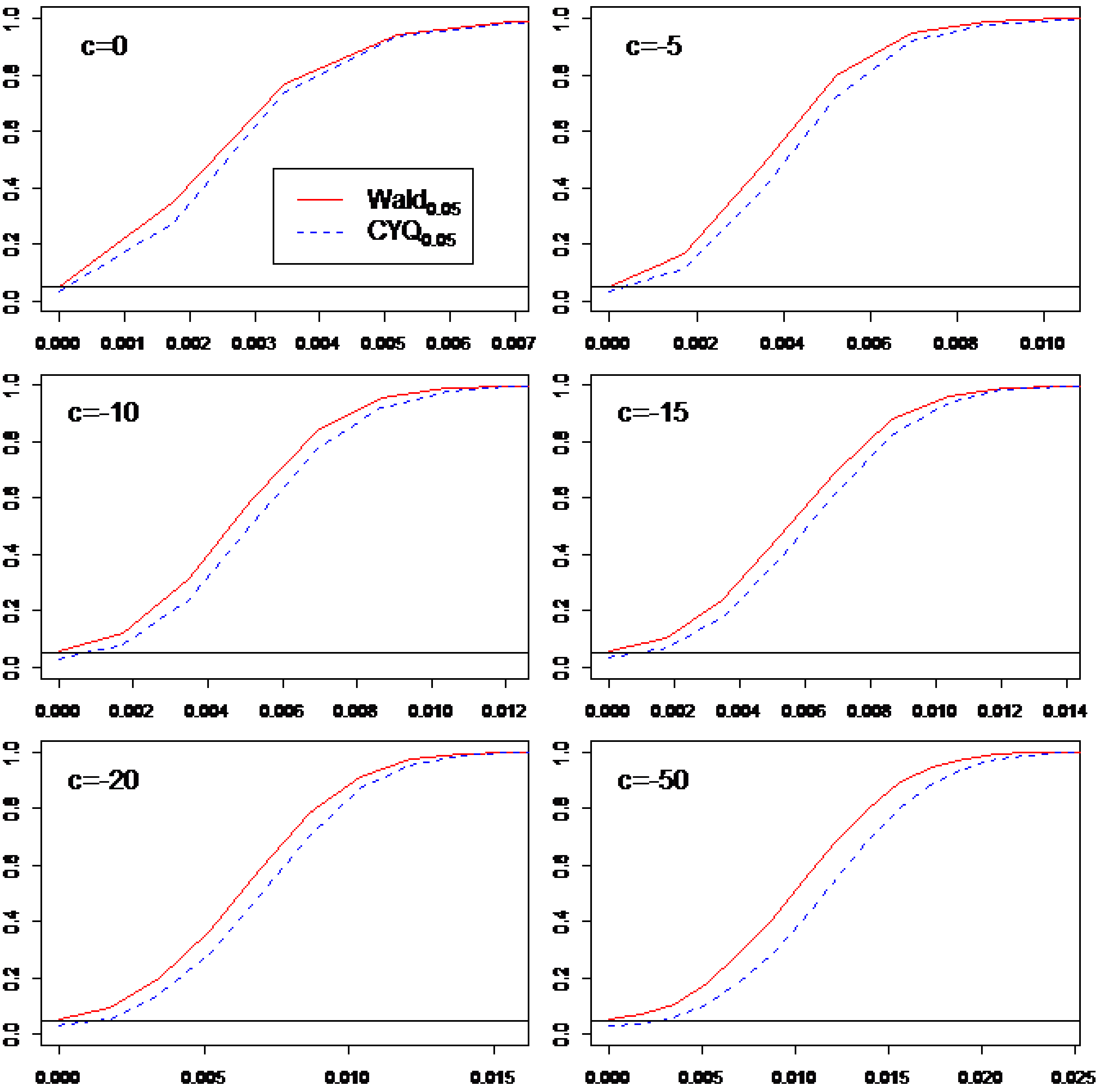


Figure A9

Power plots for sample size n=1,000, residuals' correlation coefficient $\delta = 0$ and autocorrelation coefficient $\phi = 0.5$ in the residuals of the autoregression

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of *A* increases. The solid curve ($Wald_{0.05}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line). The dashed curve ($CYQ_{0.05}$) illustrates the rejection rate using the 95% confidence interval of the Campbell and Yogo (2006) *Q*-test. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 of the main body of the study with 10,000 repetitions for a sample size of n=1,000, correlation coefficient between the residuals of regressions (22) and (23) $\delta = 0$ and autocorrelation coefficient $\phi = 0.5$ in the residuals of the autoregression (23).

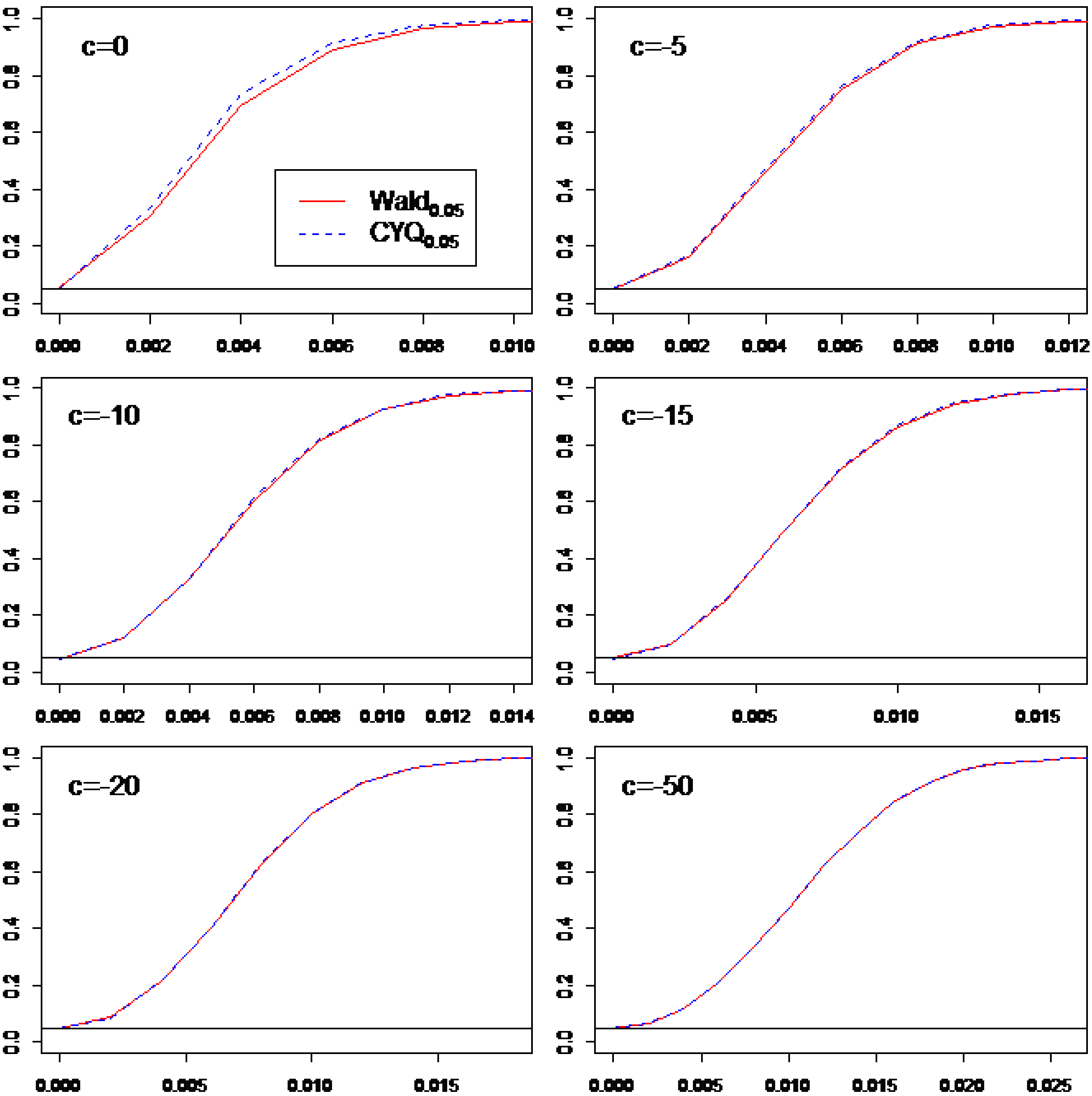


Figure A10

Power plots for sample size n=250 and residuals' correlation coefficient $\delta = -0.95$ using different kernels for the estimation of the long−run covariance matrix

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of A increases. The solid curve ($\text{Wald}_{0.05,\ \text{Bartlett}}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line), when the Bartlett kernel is used to estimate the long-run covariance matrix (à la Newey-West). The dashed curve ($\text{Wald}_{0.05,\ \text{Parzen}}$) shows the corresponding rejection rate using the Parzen kernel, while the dotted curve ($\text{Wald}_{0.05,\ \text{QS}}$) shows the corresponding rejection rate using the Quadratic Spectral kernel. Each panel corresponds to a different local-to-unity parameter C= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 of the main body of the study with 10,000 repetitions for a sample size of n=250, correlation coefficient between the residuals of regressions (22) and (23) $\delta = -0.95$ and autocorrelation coefficient $\phi = 0$ in the residuals of the autoregression (23).

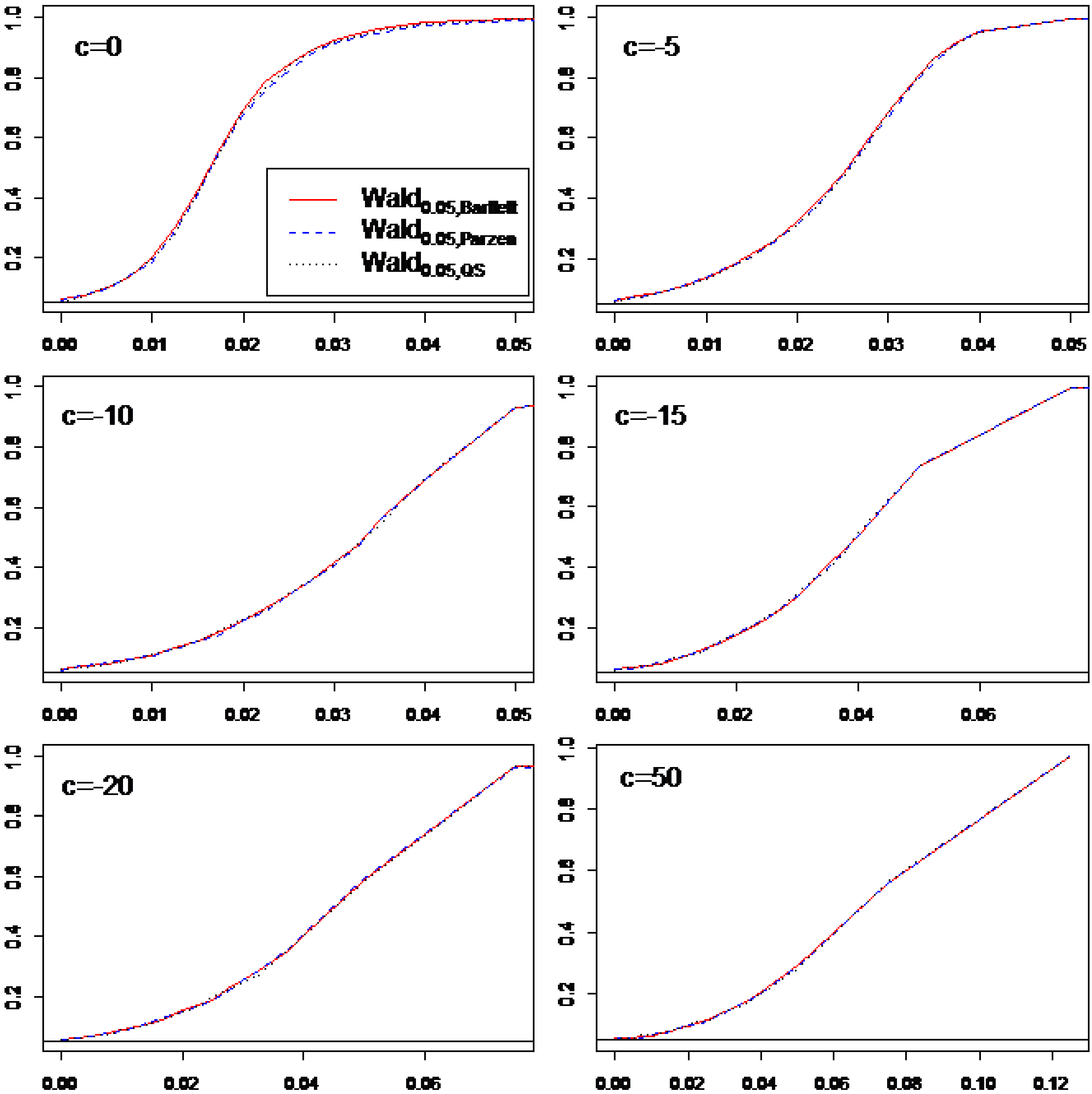


Figure A11

Power plots for sample size n=250 and residuals' correlation coefficient $\delta = -0.5$ using different kernels for the estimation of the long-run covariance matrix

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of *A* increases. The solid curve ($Wald_{0.05,\ Bartlett}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line), when the Bartlett kernel is used to estimate the long-run covariance matrix (à la Newey-West). The dashed curve ($Wald_{0.05,\ Parzen}$) shows the corresponding rejection rate using the Parzen kernel, while the dotted curve ($Wald_{0.05,\ QS}$) shows the corresponding rejection rate using the Quadratic Spectral kernel. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 of the main body of the study with 10,000 repetitions for a sample size of n=250, correlation coefficient between the residuals of regressions (22) and (23) $\delta = -0.5$ and autocorrelation coefficient $\phi = 0$ in the residuals of the autoregression (23).

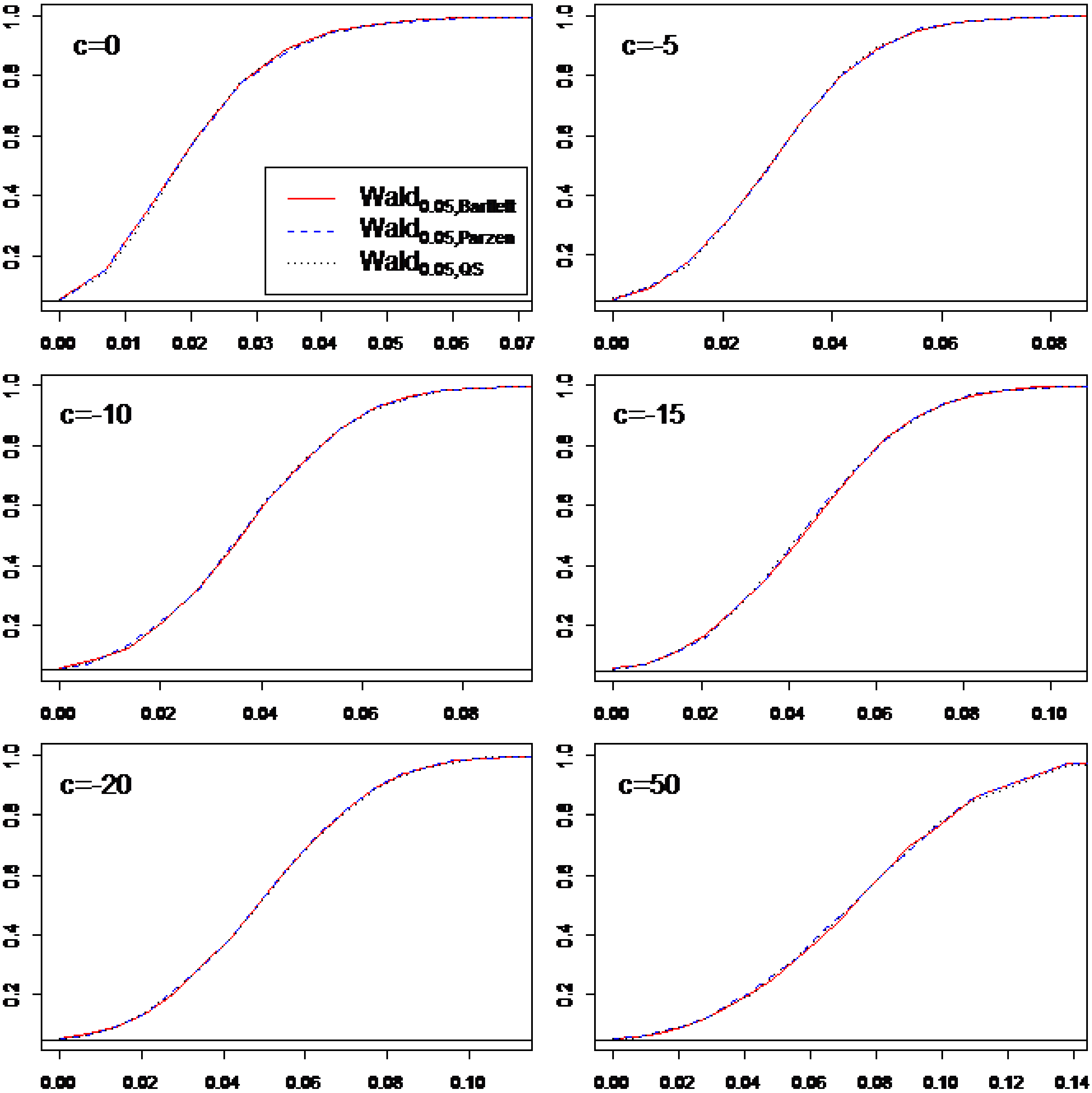


Figure A12

Power plots for sample size n=250 and residuals' correlation coefficient $\delta = 0$ using different kernels for the estimation of the long-run covariance matrix

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of A increases. The solid curve ($\text{Wald}_{0.05,\ \text{Bartlett}}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line), when the Bartlett kernel is used to estimate the long-run covariance matrix (à la Newey-West). The dashed curve ($\text{Wald}_{0.05,\ \text{Parzen}}$) shows the corresponding rejection rate using the Parzen kernel, while the dotted curve ($\text{Wald}_{0.05,\ \text{QS}}$) shows the corresponding rejection rate using the Quadratic Spectral kernel. Each panel corresponds to a different local-to-unity parameter C= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 of the main body of the study with 10,000 repetitions for a sample size of n=250, correlation coefficient between the residuals of regressions (22) and (23) $\delta = 0$ and autocorrelation coefficient $\phi = 0$ in the residuals of the autoregression (23).

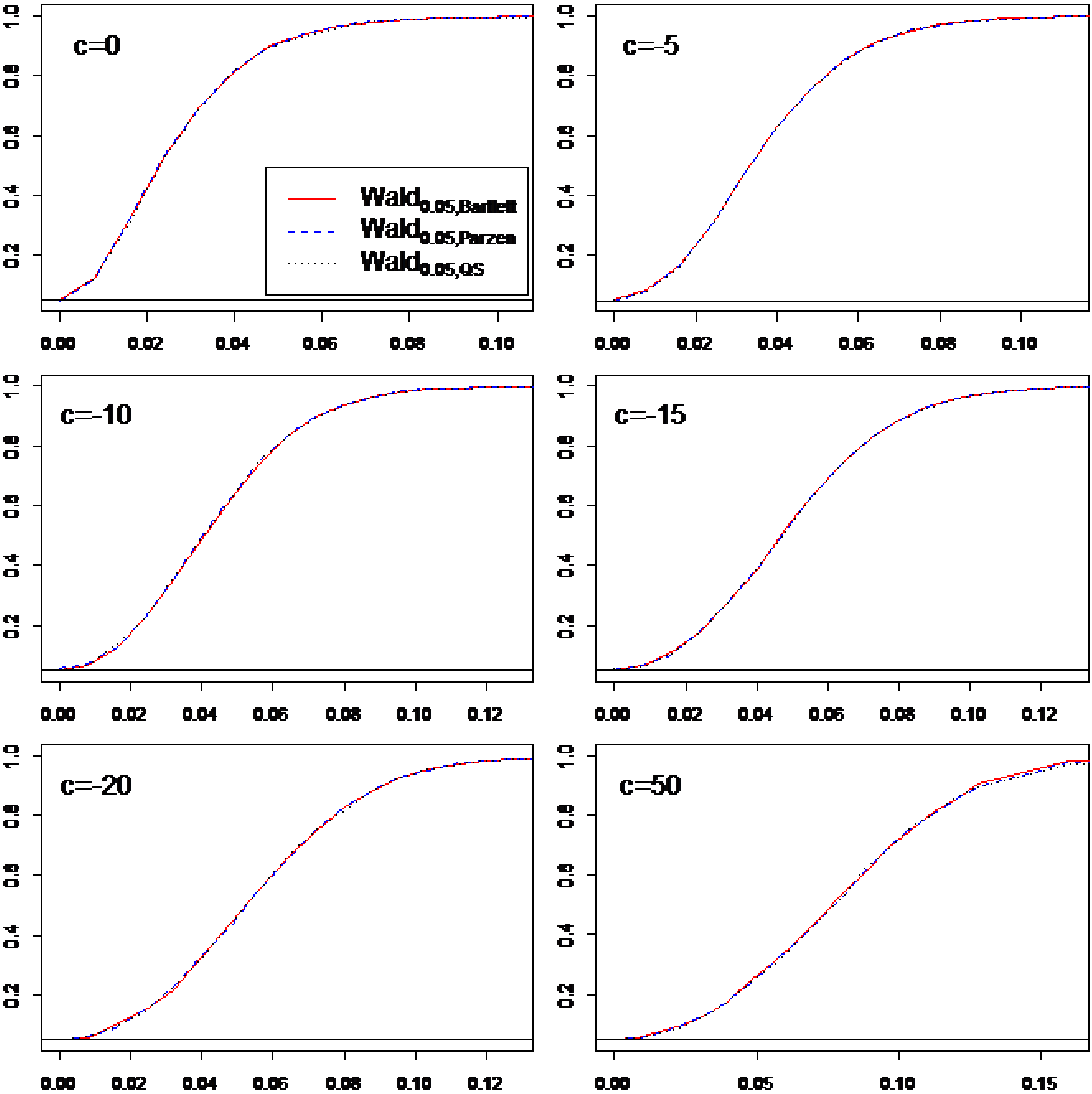


Figure A13

Power plots for sample size n=250 and residuals' correlation coefficient $\delta = -0.95$ using different number of lags in the Bartlett kernel for the estimation of the long-run covariance matrix

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of *A* increases. The solid curve ($Wald_{0.05,\ a=1/4}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line), when the lag length in the Bartlett kernel used to estimate the long-run covariance matrix is equal to $n^{1/4}$. The dashed curve ($Wald_{0.05,\ a=1/3}$) shows the corresponding rejection rate when the lag length is equal to $n^{1/3}$, while the dotted curve ($Wald_{0.05,\ a=1/2}$) shows the corresponding rejection rate when the lag length is equal to $n^{1/2}$. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 of the main body of the study with 10,000 repetitions for a sample size of n=250, correlation coefficient between the residuals of regressions (22) and (23) $\delta = -0.95$ and autocorrelation coefficient $\phi = 0$ in the residuals of the autoregression (23).

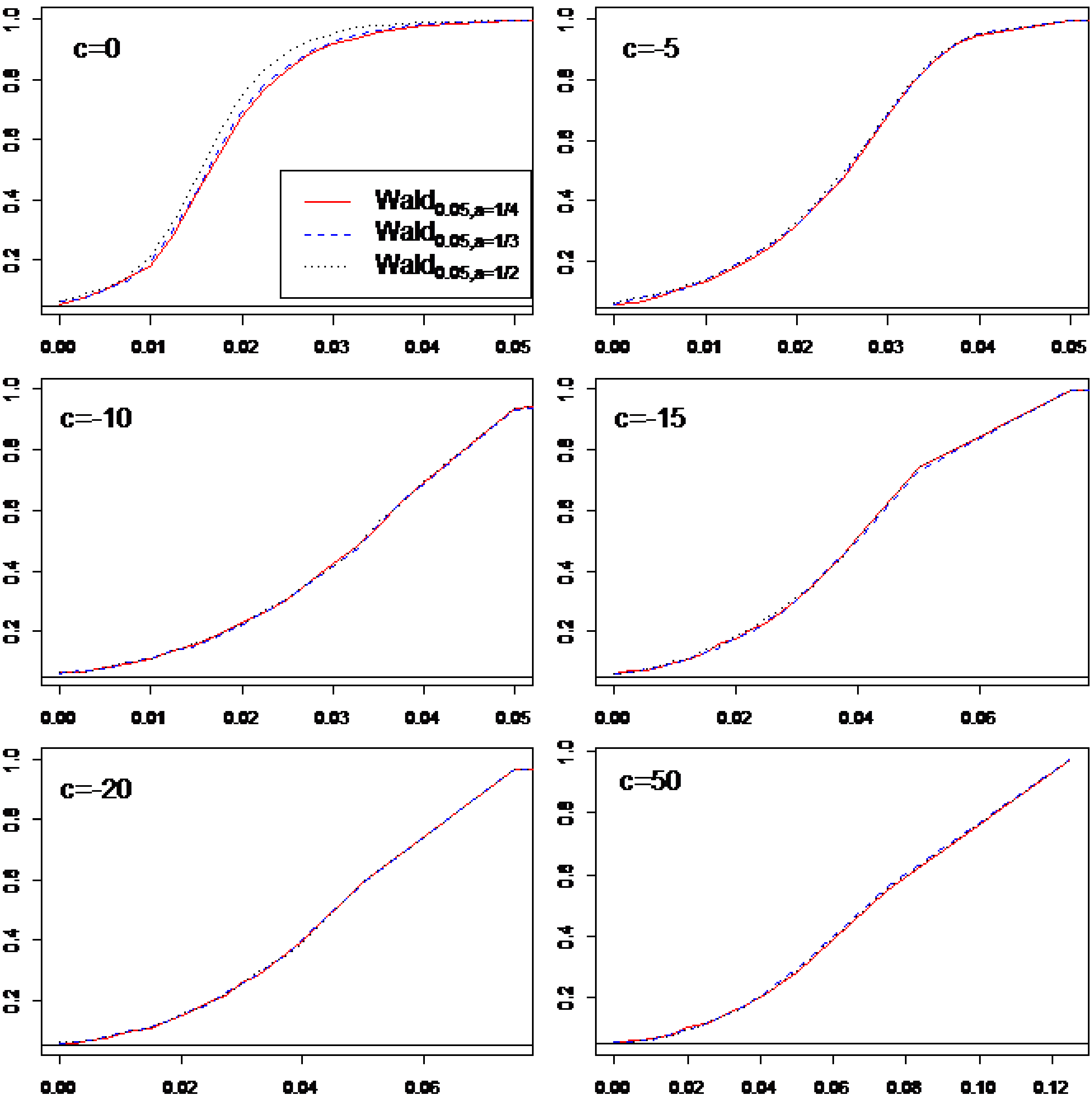


Figure A14

Power plots for sample size n=250 and residuals' correlation coefficient $\delta = -0.5$ using different number of lags in the Bartlett kernel for the estimation of the long-run covariance matrix

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of *A* increases. The solid curve ($Wald_{0.05,\ a=1/4}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line), when the lag length in the Bartlett kernel used to estimate the long-run covariance matrix is equal to $n^{1/4}$. The dashed curve ($Wald_{0.05,\ a=1/3}$) shows the corresponding rejection rate when the lag length is equal to $n^{1/3}$, while the dotted curve ($Wald_{0.05,\ a=1/2}$) shows the corresponding rejection rate when the lag length is equal to $n^{1/2}$. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 of the main body of the study with 10,000 repetitions for a sample size of n=250, correlation coefficient between the residuals of regressions (22) and (23) $\delta = -0.5$ and autocorrelation coefficient $\phi = 0$ in the residuals of the autoregression (23).

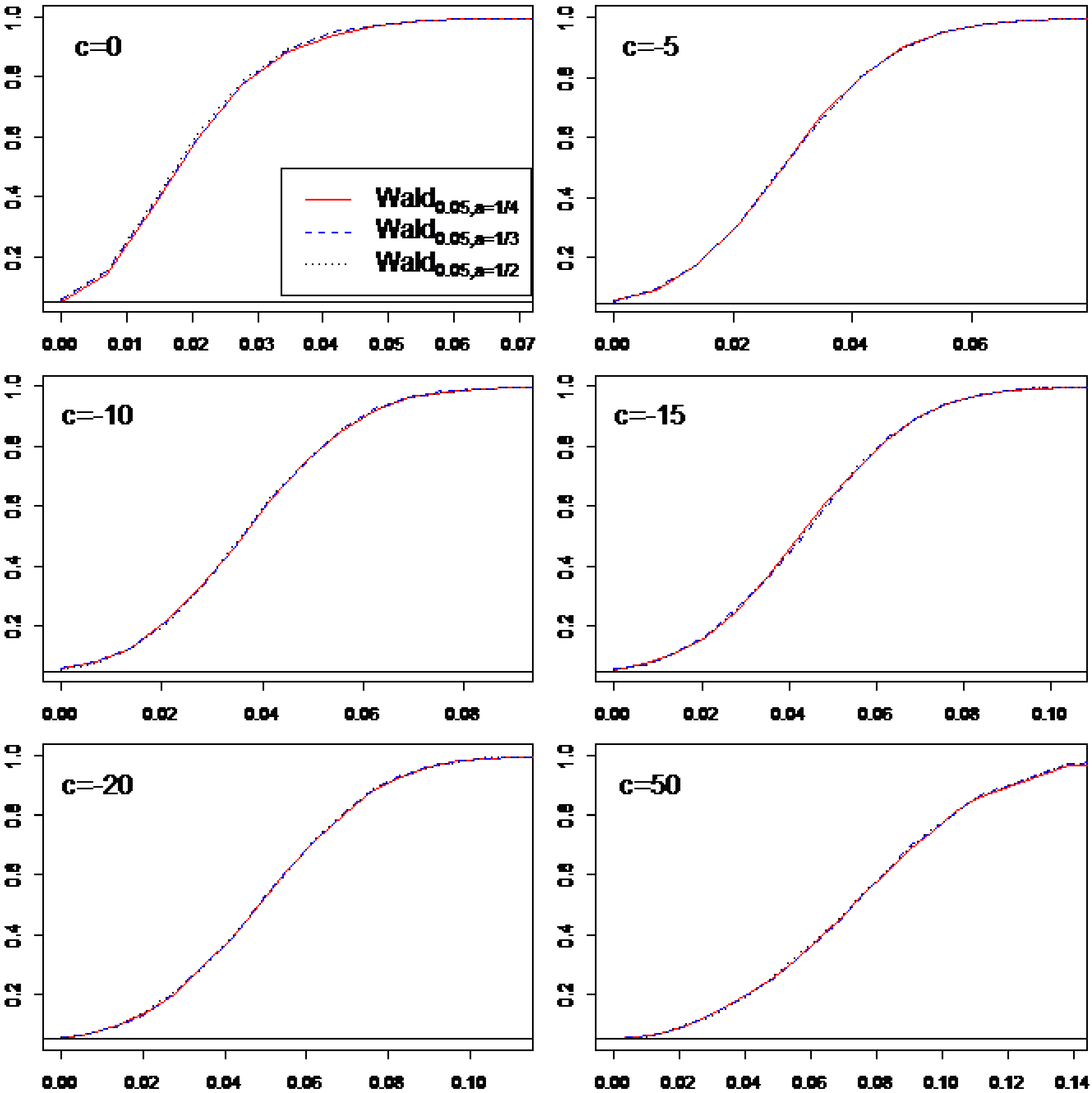


Figure A15

Power plots for sample size n=250 and residuals' correlation coefficient $\delta = 0$ using different number of lags in the Bartlett kernel for the estimation of the long-run covariance matrix

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) as the true value of *A* increases. The solid curve ($\text{Wald}_{0.05,\ a=1/4}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line), when the lag length in the Bartlett kernel used to estimate the long-run covariance matrix is equal to $n^{1/4}$. The dashed curve ($\text{Wald}_{0.05,\ a=1/3}$) shows the corresponding rejection rate when the lag length is equal to $n^{1/3}$, while the dotted curve ($\text{Wald}_{0.05,\ a=1/2}$) shows the corresponding rejection rate when the lag length is equal to $n^{1/2}$. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 of the main body of the study with 10,000 repetitions for a sample size of n=250, correlation coefficient between the residuals of regressions (22) and (23) $\delta = 0$ and autocorrelation coefficient $\phi = 0$ in the residuals of the autoregression (23).

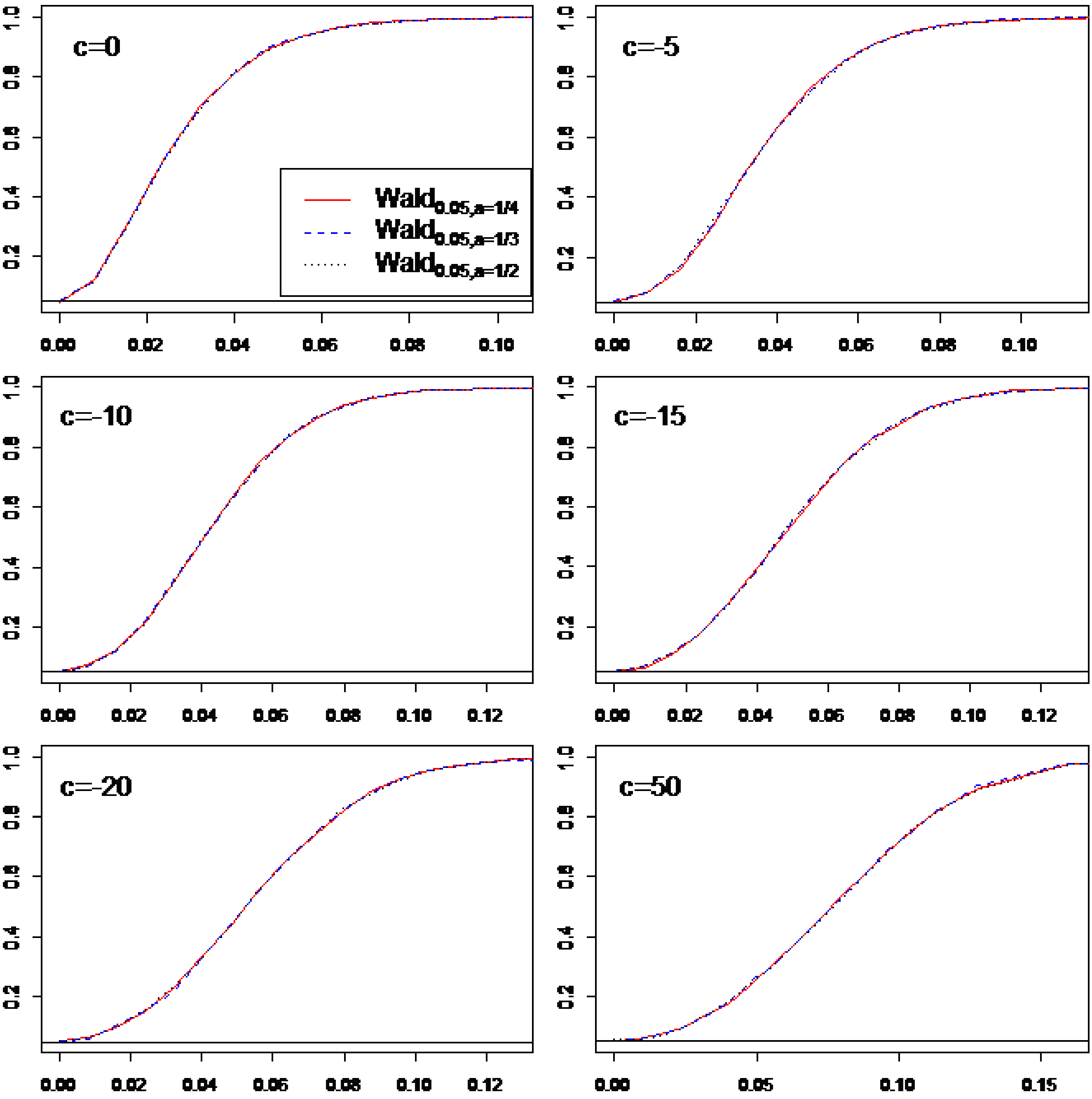


Figure A16

Rejection rates for sample size n=500 and residuals' correlation coefficient $\delta = -0.95$ using different values of β for the construction of the instrumental variable

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) using different values of β for the construction of the instrumental variable defined in (5). The solid line ($\text{Wald}_{0.05,\ A=0}$) presents the rejection rates we get using the Wald test defined in equation (19) with 5% nominal size, when the true value of *A* is zero, corresponding to the size of the test. The dotted curve ($\text{Wald}_{0.05,\ A=0.02}$) presents the corresponding rejection rates when *A*=0.02, the dashed curve ($\text{Wald}_{0.05,\ A=0.04}$) presents the corresponding rejection rates when *A*=0.04 and the dash-dot curve ($\text{Wald}_{0.05,\ A=0.06}$) presents the corresponding rejection rates when *A*=0.06. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 of the main body of the study with 10,000 repetitions for a sample size of n=500, correlation coefficient between the residuals of regressions (22) and (23) $\delta = -0.95$ and autocorrelation coefficient $\phi = 0$ in the residuals of the autoregression (23).

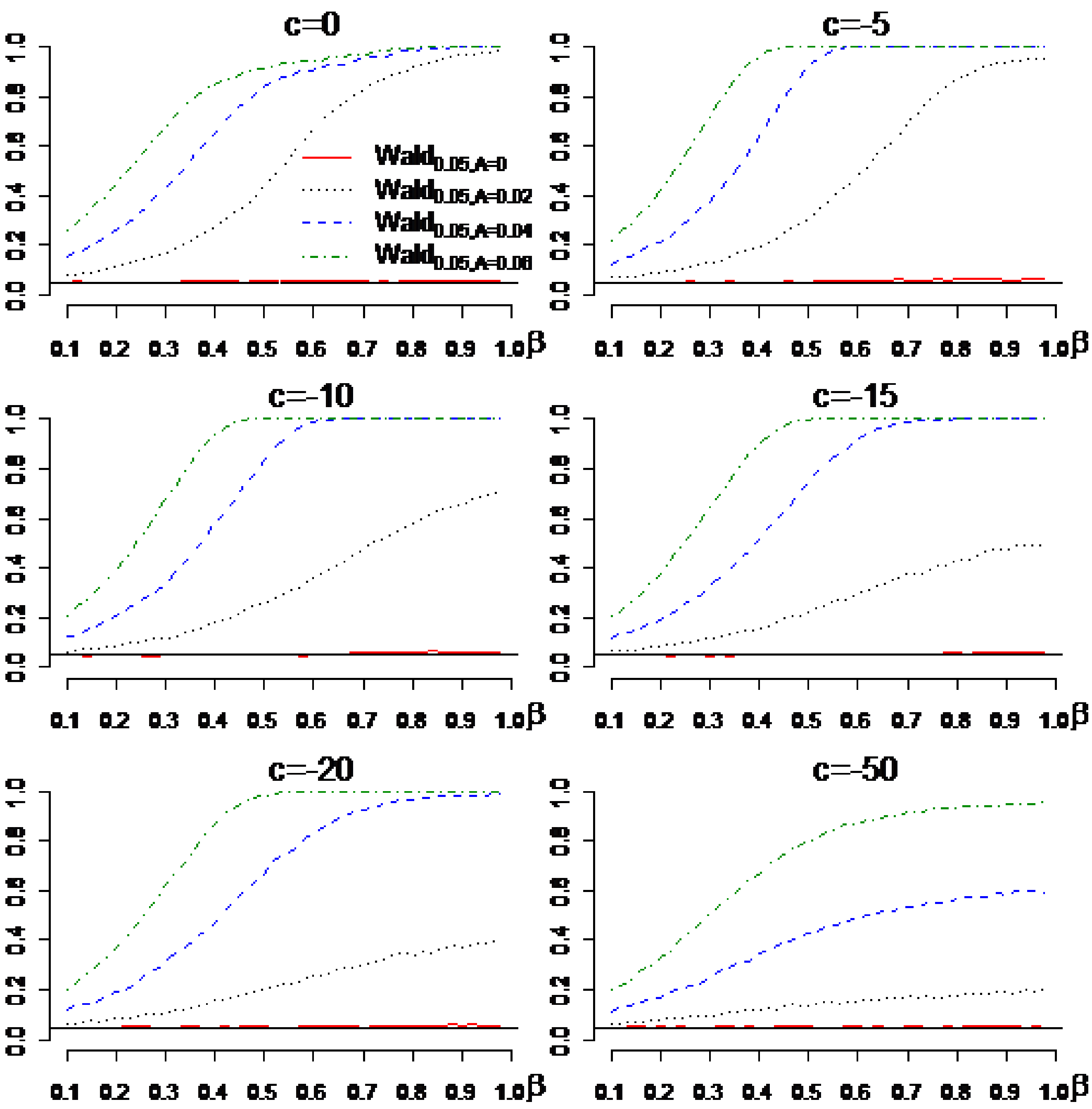


Figure A17

Rejection rates for sample size n=500 and residuals' correlation coefficient $\delta = -0.5$ using different values of β for the construction of the instrumental variable

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) using different values of β for the construction of the instrumental variable defined in (5). The solid line ($\text{Wald}_{0.05,\ A=0}$) presents the rejection rates we get using the Wald test defined in equation (19) with 5% nominal size, when the true value of *A* is zero, corresponding to the size of the test. The dotted curve ($\text{Wald}_{0.05,\ A=0.02}$) presents the corresponding rejection rates when *A*=0.02, the dashed curve ($\text{Wald}_{0.05,\ A=0.04}$) presents the corresponding rejection rates when *A*=0.04 and the dash-dot curve ($\text{Wald}_{0.05,\ A=0.06}$) presents the corresponding rejection rates when *A*=0.06. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 of the main body of the study with 10,000 repetitions for a sample size of n=500, correlation coefficient between the residuals of regressions (22) and (23) $\delta = -0.5$ and autocorrelation coefficient $\phi = 0$ in the residuals of the autoregression (23).

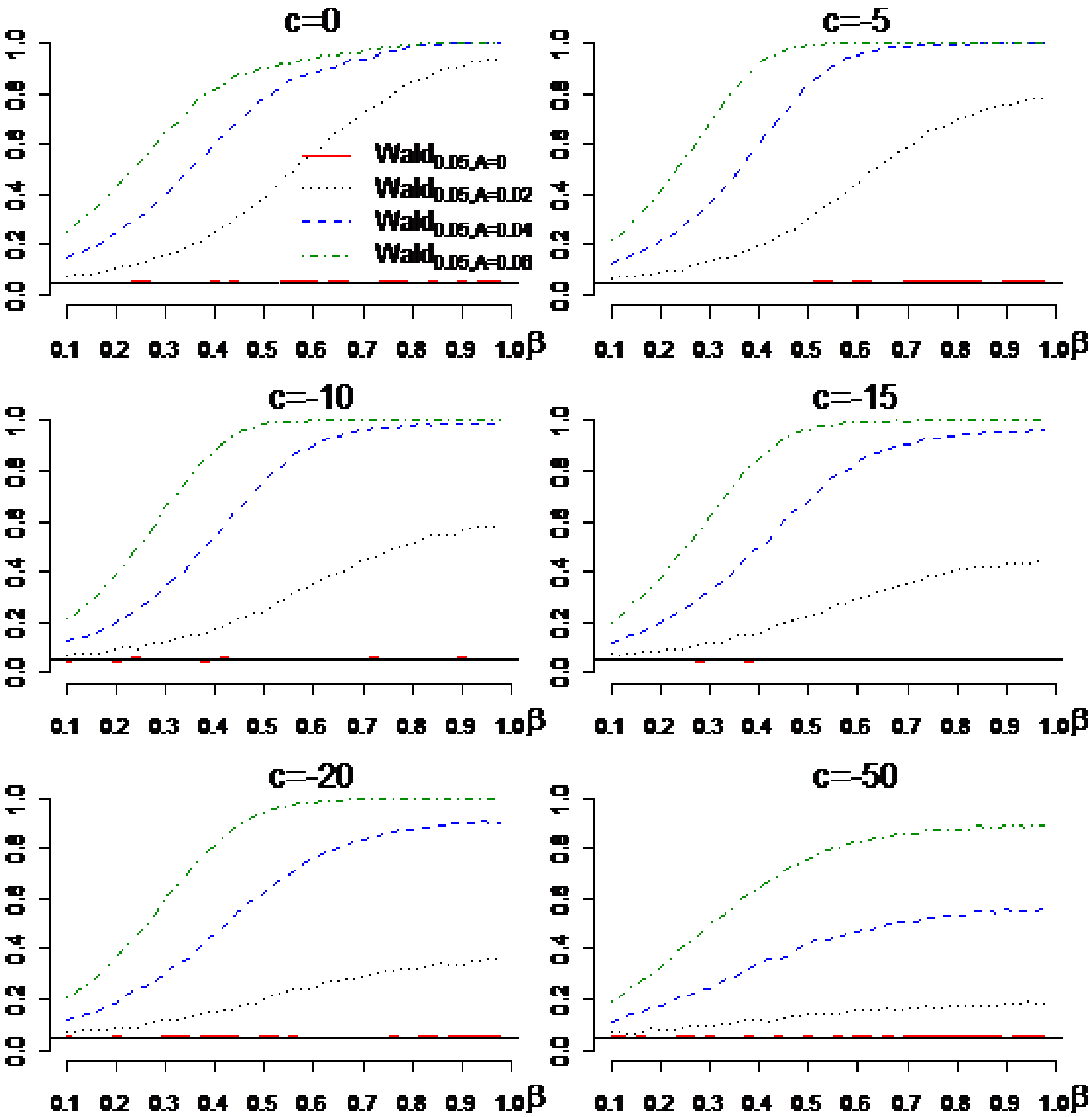


Figure A18

Rejection rates for sample size n=500 and residuals' correlation coefficient $\delta = 0$ using different values of β for the construction of the instrumental variable

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (22) using different values of β for the construction of the instrumental variable defined in (5). The solid line ($\text{Wald}_{0.05,\, A=0}$) presents the rejection rates we get using the Wald test defined in equation (19) with 5% nominal size, when the true value of *A* is zero, corresponding to the size of the test. The dotted curve ($\text{Wald}_{0.05,\, A=0.02}$) presents the corresponding rejection rates when *A*=0.02, the dashed curve ($\text{Wald}_{0.05,\, A=0.04}$) presents the corresponding rejection rates when *A*=0.04 and the dash-dot curve ($\text{Wald}_{0.05,\, A=0.06}$) presents the corresponding rejection rates when *A*=0.06. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.1 of the main body of the study with 10,000 repetitions for a sample size of n=500, correlation coefficient between the residuals of regressions (22) and (23) $\delta = 0$ and autocorrelation coefficient $\phi = 0$ in the residuals of the autoregression (23).

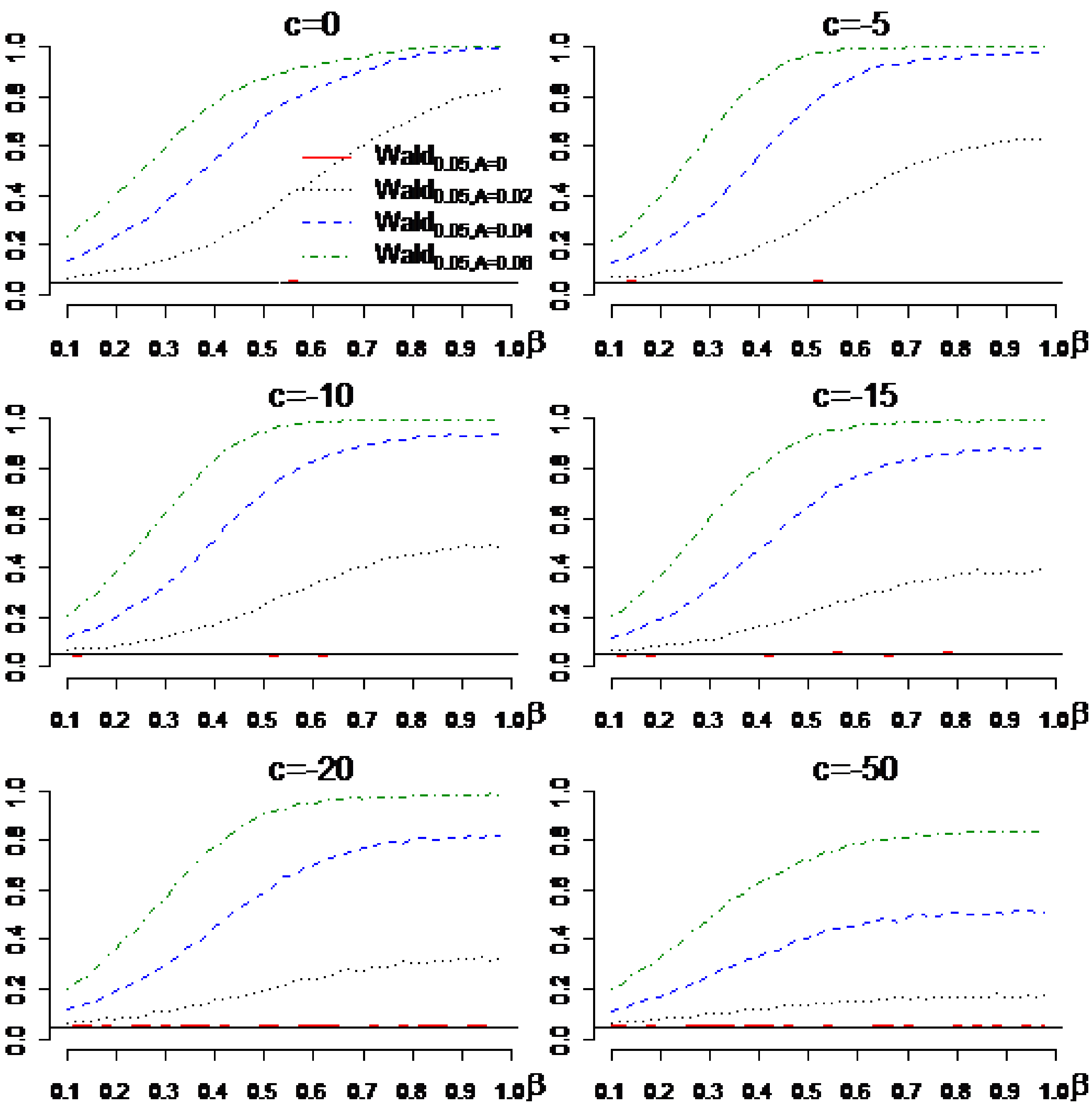


Figure A19

Power plots using a conditionally heteroskedastic DGP for stock returns, for sample size n=1,000 and residuals' correlation coefficient $\delta = -0.95$

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (34) of the Online Appendix as the true value of *A* increases, when the DGP for stock returns is conditionally heteroskedastic. In particular, the residuals of the predictive regression for stock returns are conditionally heteroskedastic, following a GARCH (1,1) process. The employed parameter values are derived from fitting a GARCH (1,1) to the residuals estimated from regressing S&P 500 value-weighted log excess returns on dividend yield, using monthly data for the period 1927−2012. The solid curve ($\text{Wald}_{0.05}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line). The dashed curve ($\text{CYQ}_{0.05}$) illustrates the rejection rate using the 95% confidence interval of the Campbell and Yogo (2006) *Q*-test. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.7 of the Online Appendix with 10,000 repetitions for a sample size of n=1,000, correlation coefficient between the residuals $\delta = -0.95$ and no autocorrelation ($\phi = 0$) in the residuals of the autoregression.

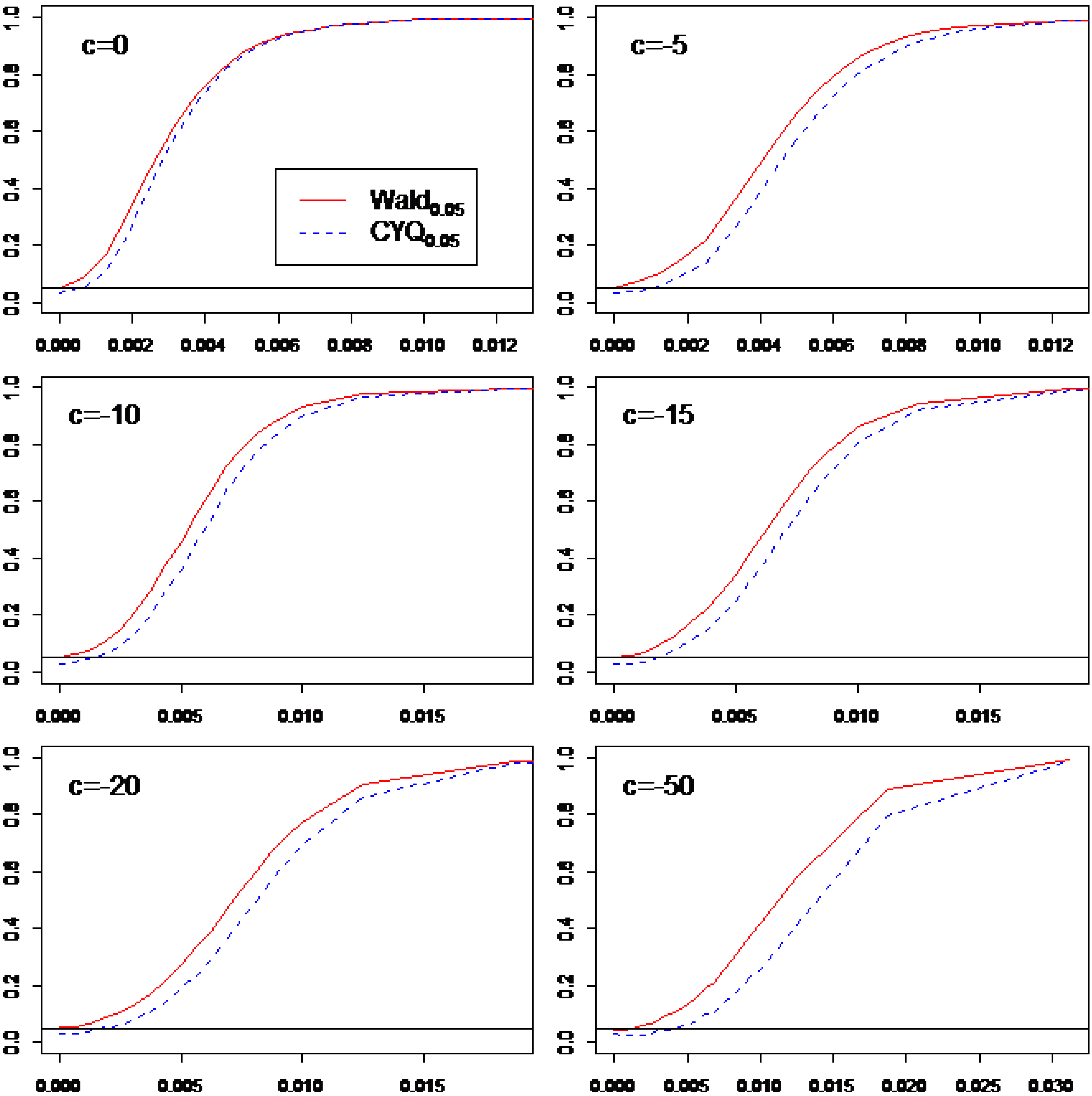


Figure A20

Power plots using a conditionally heteroskedastic DGP for stock returns, for sample size n=1,000 and residuals' correlation coefficient $\delta = -0.5$

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (34) of the Online Appendix as the true value of *A* increases, when the DGP for stock returns is conditionally heteroskedastic. In particular, the residuals of the predictive regression for stock returns are conditionally heteroskedastic, following a GARCH (1,1) process. The employed parameter values are derived from fitting a GARCH (1,1) process to the residuals estimated from regressing S&P 500 value-weighted log excess returns on dividend yield, using monthly data for the period 1927−2012. The solid curve ($Wald_{0.05}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line). The dashed curve ($CYQ_{0.05}$) illustrates the rejection rate using the 95% confidence interval of the Campbell and Yogo (2006) *Q*-test. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.7 of the Online Appendix with 10,000 repetitions for a sample size of n=1,000, correlation coefficient between the residuals $\delta = -0.5$ and no autocorrelation ($\phi = 0$) in the residuals of the autoregression.

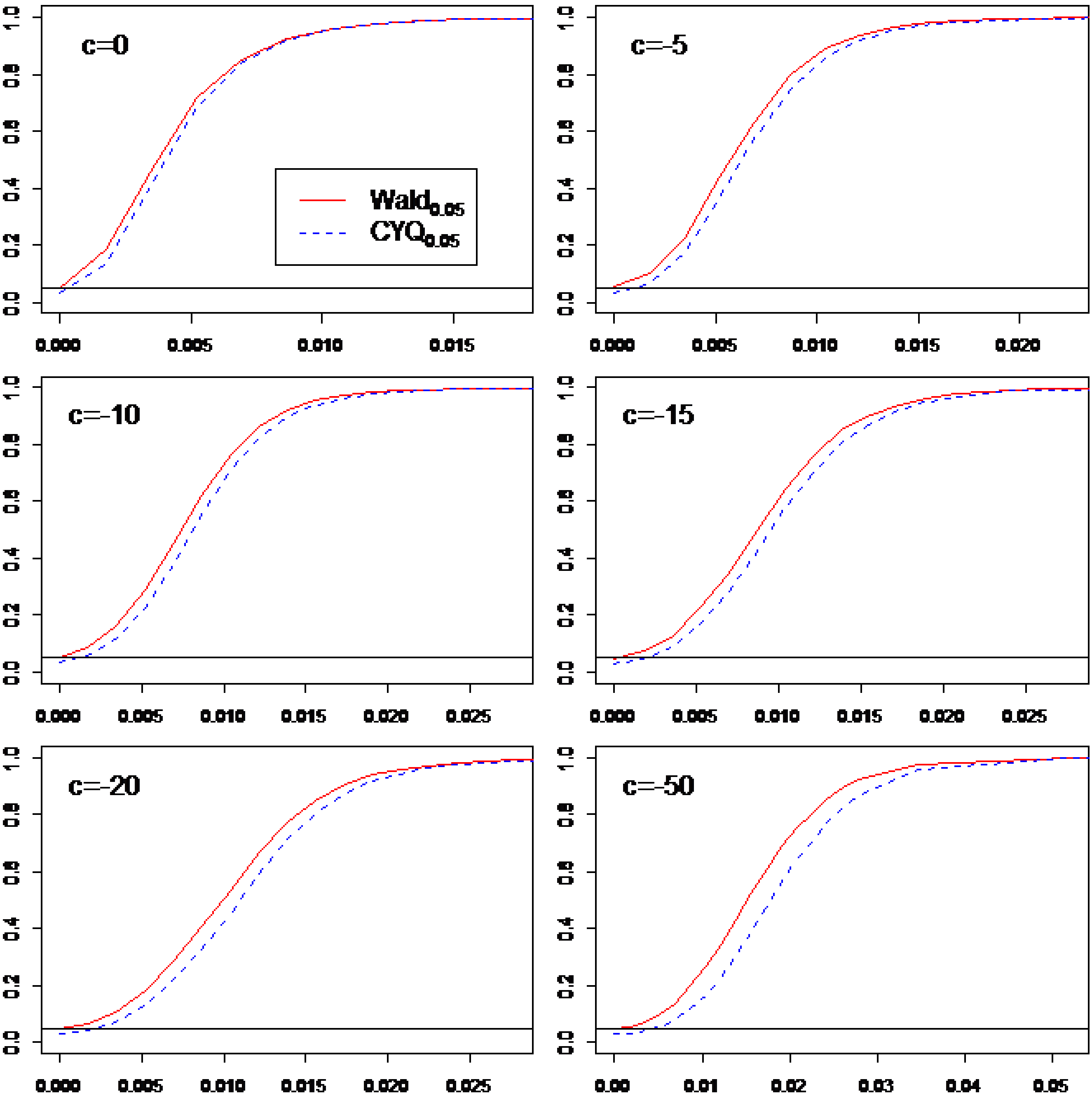


Figure A21

Power plots using a conditionally heteroskedastic DGP for stock returns, for sample size n=1,000 and residuals' correlation coefficient $\delta = 0$

This figure shows the rejection rates for tests of the null hypothesis $H_0 : A = 0$ versus the alternative $H_1 : A \neq 0$ in (34) of the Online Appendix as the true value of *A* increases, when the DGP for stock returns is conditionally heteroskedastic. In particular, the residuals of the predictive regression for stock returns are conditionally heteroskedastic, following a GARCH (1,1) process. The employed parameter values are derived from fitting a GARCH (1,1) process to the residuals estimated from regressing S&P 500 value-weighted log excess returns on dividend yield, using monthly data for the period 1927−2012. The solid curve ($Wald_{0.05}$) illustrates the rejection rate we get using the Wald test, defined in equation (19), with 5% nominal size (horizontal line). The dashed curve ($CYQ_{0.05}$) illustrates the rejection rate using the 95% confidence interval of the Campbell and Yogo (2006) *Q*-test. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 2.7 of the Online Appendix with 10,000 repetitions for a sample size of n=1,000, correlation coefficient between the residuals $\delta = 0$ and no autocorrelation ($\phi = 0$) in the residuals of the autoregression.

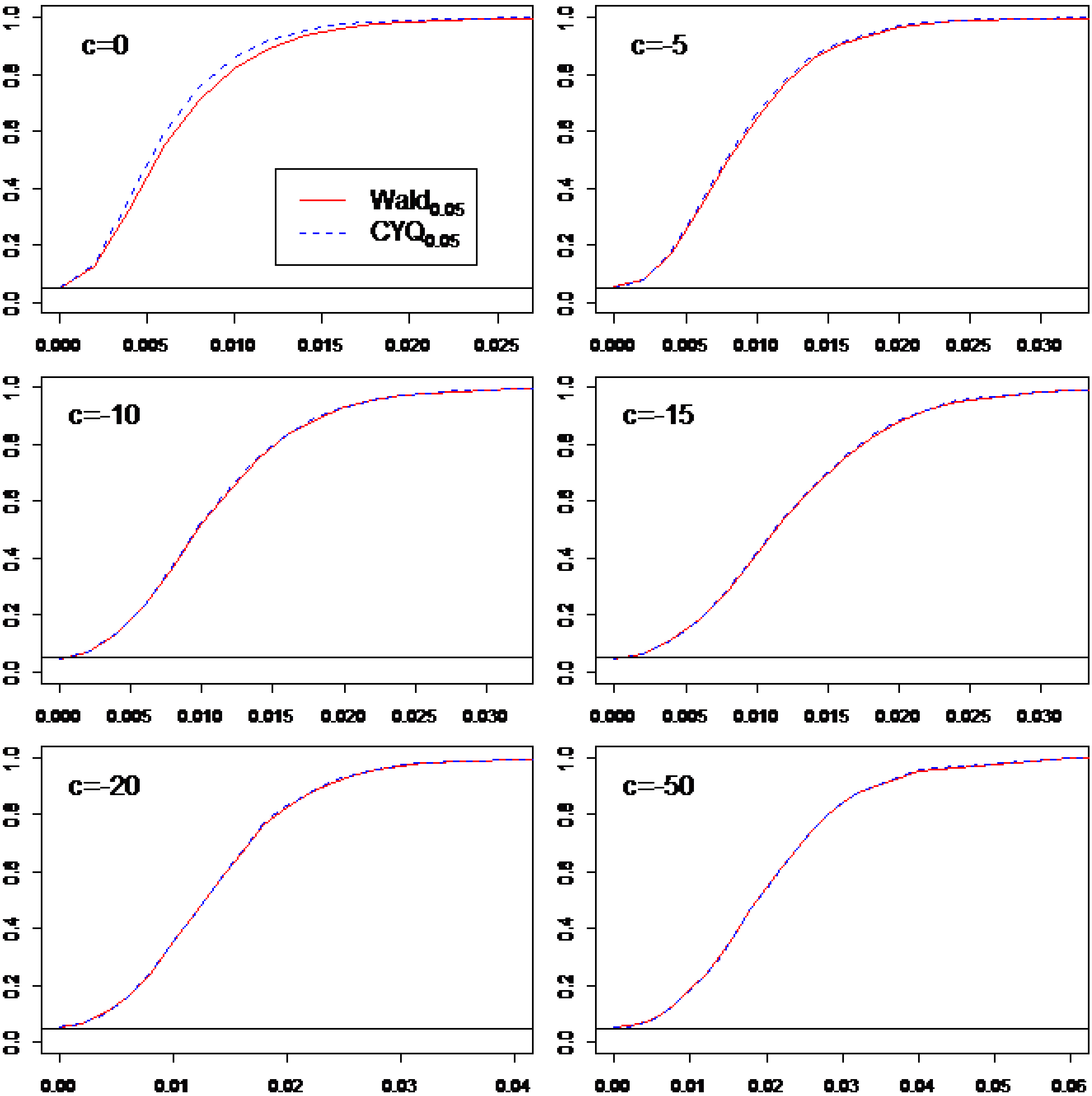


Figure A22

Power plots for long−horizon Wald test, sample size n=1,000 and residuals' correlation coefficient δ =−0.95

This figure shows the rejection rates, derived from *K*-horizon univariate predictive regressions as in equation (30), for tests of the null hypothesis $H_0 : A = 0$ in the DGP (22), as the true value of *A* increases. The solid curve ($Wald_{0.05,\ K=12}$) illustrates the rejection rate we get using the long-horizon Wald test, defined in equation (34), with 5% nominal size (horizontal line), when the predictive horizon is *K*=12. The dotted curve ($Wald_{0.05,\ K=36}$) illustrates the corresponding rejection rate when the horizon is *K*=36. The dashed curve ($Wald_{0.05,\ K=60}$) illustrates the corresponding rejection rate when the horizon is *K*=60. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 5.2 with 10,000 repetitions for a sample size of n=1,000, correlation coefficient between the residuals of regressions (22) and (23) δ =−0.95 and no autocorrelation in the residuals of the autoregressive equation, i.e., ϕ =0 in (24).

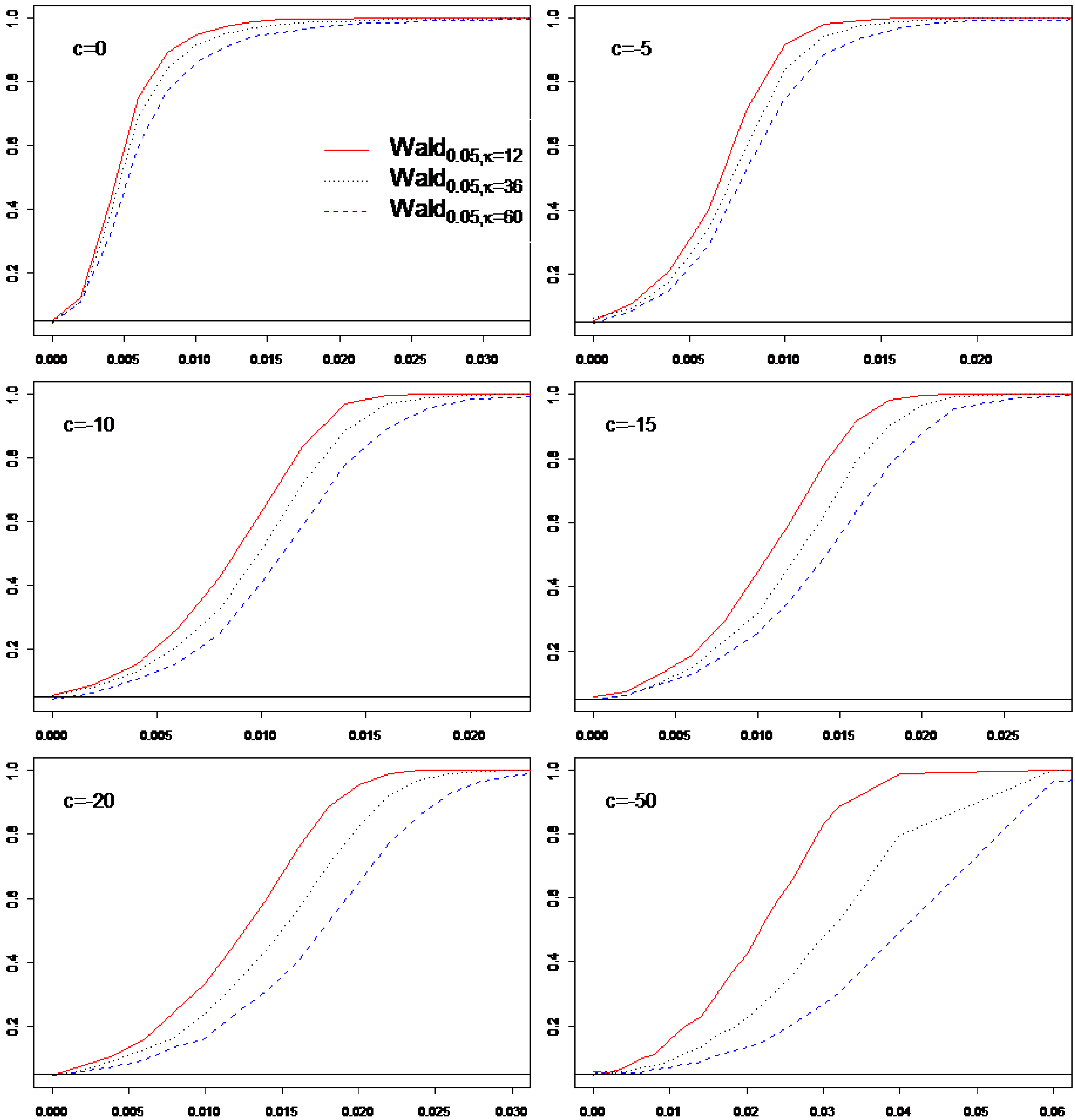


Figure A23

Power plots for long−horizon Wald test, sample size n=1,000 and residuals' correlation coefficient $\delta = -0.5$

This figure shows the rejection rates, derived from *K*-horizon univariate predictive regressions as in equation (30), for tests of the null hypothesis $H_0 : A = 0$ in the DGP (22), as the true value of *A* increases. The solid curve ($\text{Wald}_{0.05,\ K=12}$) illustrates the rejection rate we get using the long-horizon Wald test, defined in equation (34), with 5% nominal size (horizontal line), when the predictive horizon is *K*=12. The dotted curve ($\text{Wald}_{0.05,\ K=36}$) illustrates the corresponding rejection rate when the horizon is *K*=36. The dashed curve ($\text{Wald}_{0.05,\ K=60}$) illustrates the corresponding rejection rate when the horizon is *K*=60. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 5.2 with 10,000 repetitions for a sample size of n=1,000, correlation coefficient between the residuals of regressions (22) and (23) $\delta = -0.5$ and no autocorrelation in the residuals of the autoregressive equation, i.e., $\phi = 0$ in (24).

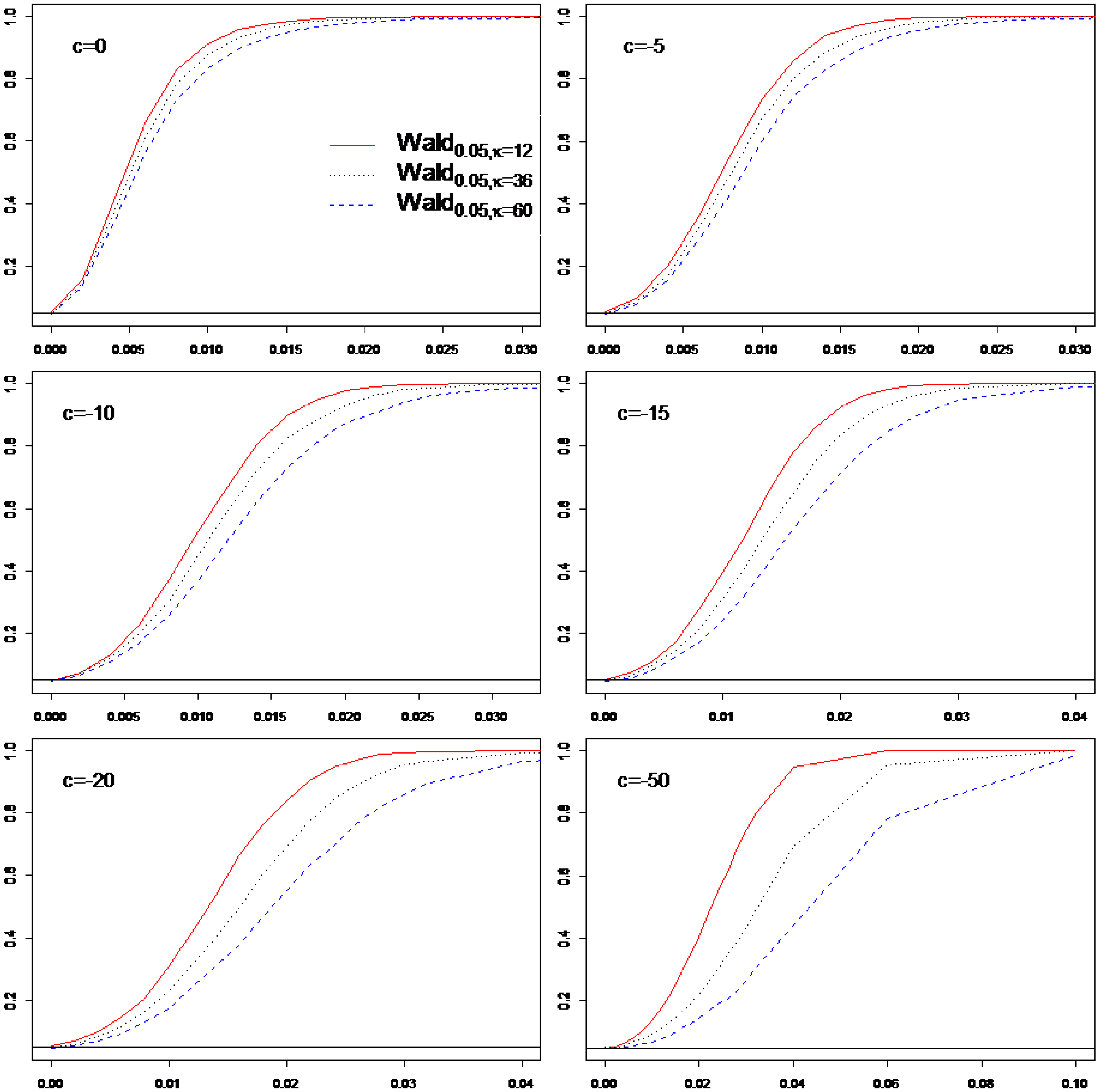


Figure A24

Power plots for long−horizon Wald test, sample size n=1,000 and residuals' correlation coefficient $\delta=0$

This figure shows the rejection rates, derived from *K*-horizon univariate predictive regressions as in equation (30), for tests of the null hypothesis $H_0 : A = 0$ in the DGP (22), as the true value of *A* increases. The solid curve ($\text{Wald}_{0.05,\ K=12}$) illustrates the rejection rate we get using the long-horizon Wald test, defined in equation (34), with 5% nominal size (horizontal line), when the predictive horizon is *K*=12. The dotted curve ($\text{Wald}_{0.05,\ K=36}$) illustrates the corresponding rejection rate when the horizon is *K*=36. The dashed curve ($\text{Wald}_{0.05,\ K=60}$) illustrates the corresponding rejection rate when the horizon is *K*=60. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 5.2 with 10,000 repetitions for a sample size of n=1,000, correlation coefficient between the residuals of regressions (22) and (23) $\delta=0$ and no autocorrelation in the residuals of the autoregressive equation, i.e., $\phi=0$ in (24).

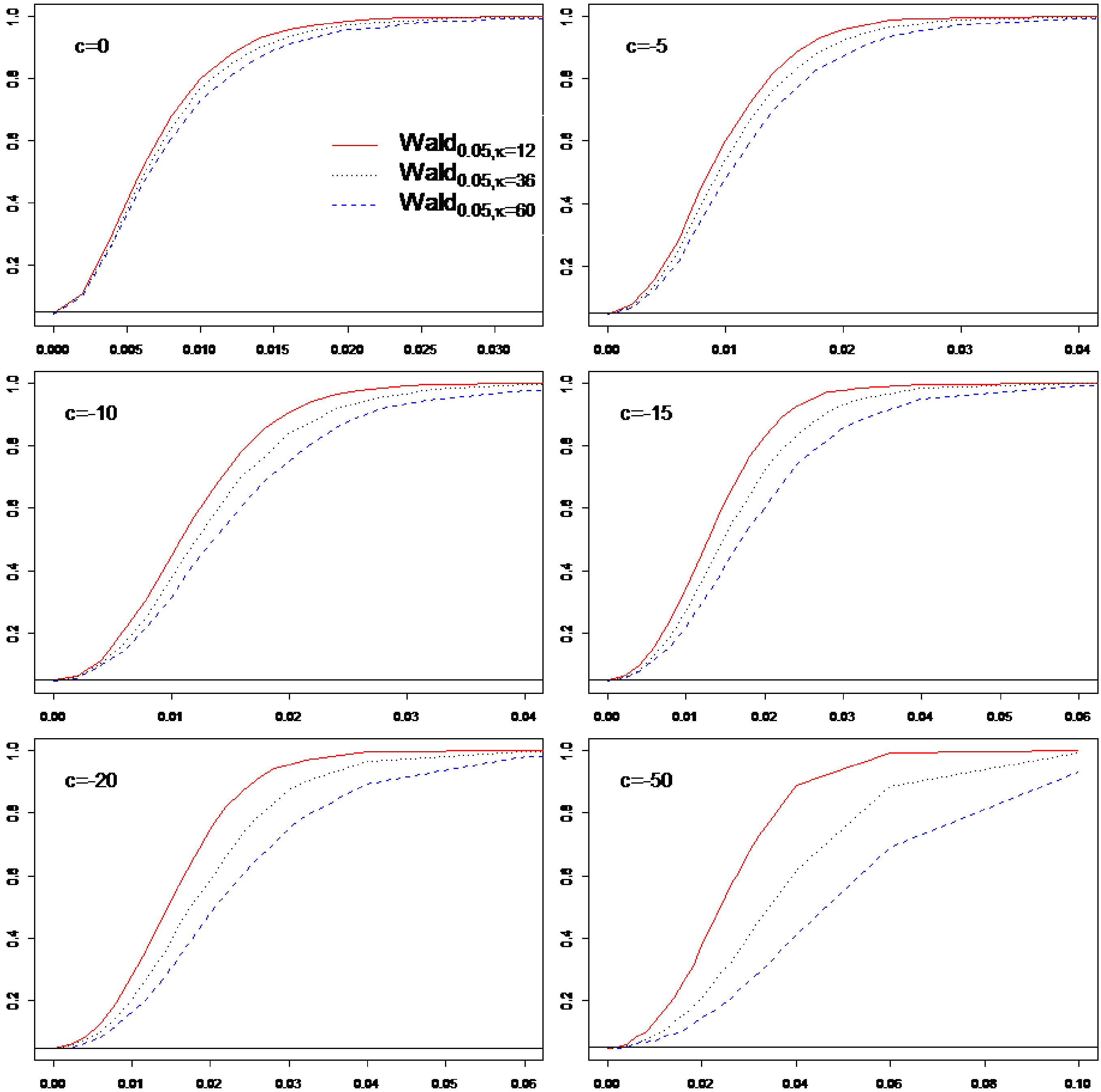


Figure A25

Power plots for long−horizon Wald test, sample size n=500 and residuals' correlation coefficient $\delta = -0.95$

This figure shows the rejection rates, derived from *K*-horizon univariate predictive regressions as in equation (30), for tests of the null hypothesis $H_0 : A = 0$ in the DGP (22), as the true value of *A* increases. The solid curve ($Wald_{0.05,\ K=4}$) illustrates the rejection rate we get using the long-horizon Wald test, defined in equation (34), with 5% nominal size (horizontal line), when the predictive horizon is *K*=4. The dotted curve ($Wald_{0.05,\ K=12}$) illustrates the corresponding rejection rate when the horizon is *K*=12. The dashed curve ($Wald_{0.05,\ K=20}$) illustrates the corresponding rejection rate when the horizon is *K*=20. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 5.2 of the main body of the study with 10,000 repetitions for a sample size of n=500, correlation coefficient between the residuals of regressions (22) and (23) in the main body of the study $\delta = -0.95$ and no autocorrelation in the residuals of the autoregressive equation, i.e., $\phi = 0$ in (24).

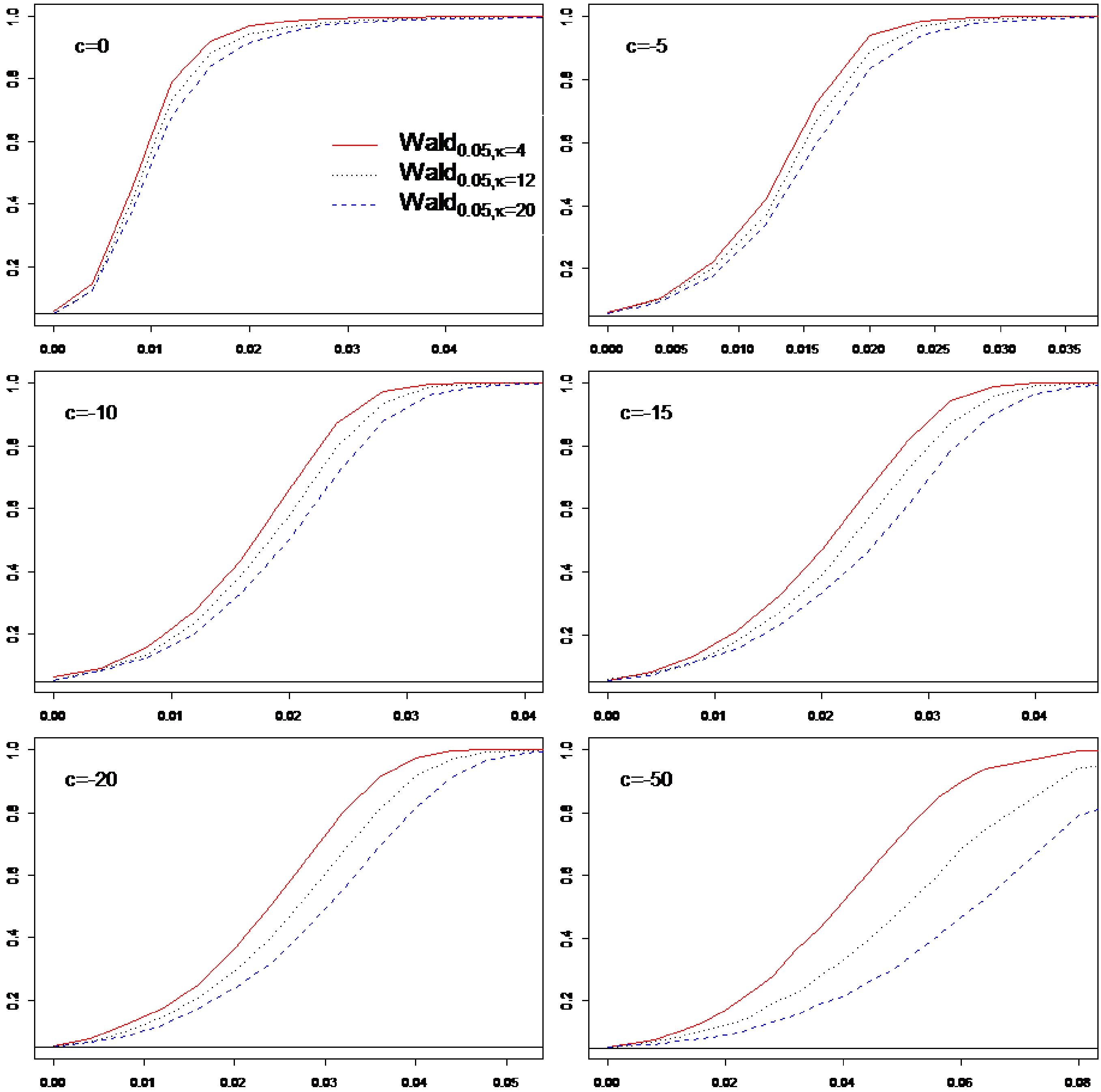


Figure A26

Power plots for long−horizon Wald test, sample size n=500 and residuals' correlation coefficient $\delta = -0.5$

This figure shows the rejection rates, derived from *K*-horizon univariate predictive regressions as in equation (30), for tests of the null hypothesis $H_0 : A = 0$ in the DGP (22), as the true value of *A* increases. The solid curve ($\text{Wald}_{0.05,\ K=4}$) illustrates the rejection rate we get using the long-horizon Wald test, defined in equation (34), with 5% nominal size (horizontal line), when the predictive horizon is *K*=4. The dotted curve ($\text{Wald}_{0.05,\ K=12}$) illustrates the corresponding rejection rate when the horizon is *K*=12. The dashed curve ($\text{Wald}_{0.05,\ K=20}$) illustrates the corresponding rejection rate when the horizon is *K*=20. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 5.2 of the main body of the study with 10,000 repetitions for a sample size of n=500, correlation coefficient between the residuals of regressions (22) and (23) in the main body of the study $\delta = -0.5$ and no autocorrelation in the residuals of the autoregressive equation, i.e., $\phi = 0$ in (24).

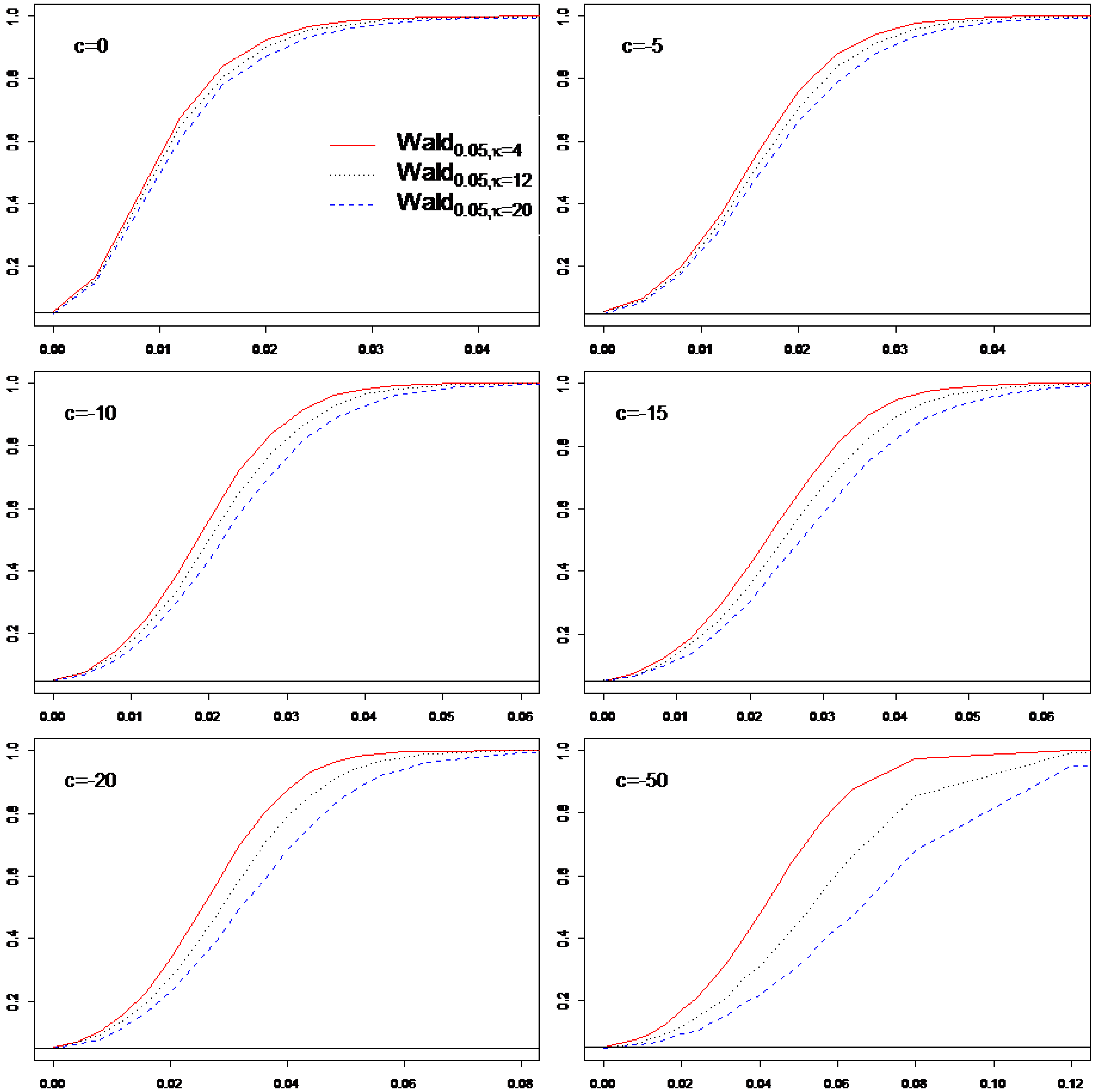


Figure A27

Power plots for long−horizon Wald test, sample size n=500 and residuals' correlation coefficient δ =0

This figure shows the rejection rates, derived from *K*-horizon univariate predictive regressions as in equation (30), for tests of the null hypothesis $H_0 : A = 0$ in the DGP (22), as the true value of *A* increases. The solid curve ($Wald_{0.05,\ K=4}$) illustrates the rejection rate we get using the long-horizon Wald test, defined in equation (34), with 5% nominal size (horizontal line), when the predictive horizon is *K*=4. The dotted curve ($Wald_{0.05,\ K=12}$) illustrates the corresponding rejection rate when the horizon is *K*=12. The dashed curve ($Wald_{0.05,\ K=20}$) illustrates the corresponding rejection rate when the horizon is *K*=20. Each panel corresponds to a different local-to-unity parameter *C*= 0, −5, −10, −15, −20 and −50. These rejection rates have been calculated using Monte Carlo simulations described in Section 5.2 of the main body of the study with 10,000 repetitions for a sample size of n=500, correlation coefficient between the residuals of regressions (22) and (23) in the main body of the study $\delta = 0$ and no autocorrelation in the residuals of the autoregressive equation, i.e., $\phi = 0$ in (24).