Essex Finance Centre
Working Paper Series

Working Paper No 67: 02-2021

"Simple Tests for Stock Return Predictability with Good Size and Power Properties"

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

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

Simple Tests for Stock Return Predictability with Good Size and Power Properties*

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

[a] School of Economics, University of Nottingham
[b] Essex Business School, University of Essex

Abstract

We develop easy-to-implement tests for return predictability which, relative to extant tests in the literature, display attractive finite sample size control and power across a wide range of persistence and endogeneity levels for the predictor. Our approach is based on the standard regression t-ratio and a variant where the predictor is quasi-GLS (rather than OLS) demeaned. In the strongly persistent near-unit root environment, the limiting null distributions of these statistics depend on the endogeneity and local-to-unity parameters characterising the predictor. Analysis of the asymptotic local power functions of feasible implementations of these two tests, based on asymptotically conservative critical values, motivates a switching procedure between the two, employing the quasi-GLS demeaned variant unless the magnitude of the estimated endogeneity correlation parameter is small. Additionally, if the data suggests the predictor is weakly persistent, our approach switches into the standard t-ratio test with reference to standard normal critical values.

Keywords: predictive regression; persistence; endogeneity; quasi-GLS demeaning; unit root test; hybrid statistic.

JEL classification: C12, C22

*We dedicate this paper to Pierre Perron and the enormous contributions he has made to the science of econometrics. All three of us have benefited hugely not only from Pierre's intellectual contributions to the discipline, but also from his professional generosity and kindness. We are grateful to two anonymous referees and the Editor, Serena Ng, for their helpful and constructive comments on previous drafts of this paper. Taylor gratefully acknowledges financial support provided by the Economic and Social Research Council of the United Kingdom under research grant ES/R00496X/1. Correspondence to: Robert Taylor, Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. Email: robert.taylor@essex.ac.uk

1 Introduction

A large body of empirical research has been undertaken investigating whether stock returns can be predicted using publicly available data. A wide range of financial and macroeconomic variables has been considered as putative predictors for returns, including: valuation ratios such as the dividend-price ratio, dividend yield, earnings-price ratio, and book-to-market ratio; various interest rates and interest rate spreads, and macroeconomic variables including inflation and industrial production; see, for example, Fama (1981), Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988a,b), Fama and French (1988, 1989) and Fama (1990).

Empirical evidence on the predictability of returns largely derives from inference obtained from predictive regressions and, as such, the size and power properties of tests from these regressions are of fundamental importance. These depend on the time series properties of the predictor used, in particular its degree of persistence and endogeneity. Data analysis presented in, among others, Campbell and Yogo (2006) [hereafter CY] and Welch and Goyal (2008), suggests that many of the variables used in predictive regressions are highly persistent with autoregressive roots close to unity, and that a strong negative correlation often exists between returns and the predictor's innovations, such that the predictive regressor is endogenous.

A number of likelihood-based predictability tests have been developed, designed to be asymptotically valid when the predictor is strongly persistent and endogenous; see, in particular, Cavanagh *et al.* (1995), Lewellen (2004), CY and Jansson and Moreira (2006). These approaches are based on a formulation where the predictor, x_{t-1} say, is assumed to follow a first-order autoregression with a local-to-unity coefficient $\phi = 1 - c/T$, where c is a finite unknown constant and T is the sample size. Of these, the Q test of CY is widely viewed as the state of the art methodology in the literature for testing the predictability of stock returns with highly persistent regressors. A major drawback with these tests, however, is that they are invalid if the predictor is weakly persistent (stationary). Alternative tests based on instrumental variable [IV] estimation have also been developed; see, among others, Phillips and Magdalinos (2009), Kostakis *et al.* (2015) and Breitung and Demetrescu (2015). Here a stochastic instrument is constructed from the predictor which, by design, is less persistent than a local-to-unity process. The IV-based tests are asymptotically valid regardless of whether the predictor is local-to-unity or weakly persistent, but their power is not as high as the likelihood-based tests when the predictor is strongly persistent. Breitung and Demetrescu (2015) therefore also propose a combined instrument test using two instruments: the first as described above, the second a trending variable independent of the predictor. This test is designed such that, in large samples, it selects the second instrument when the predictor is local-to-unity but reverts to the first instrument otherwise. A significant drawback, however, is that it can only be implemented as a two-tailed test and so if the direction of predictability is known, it can have significantly lower power than one-sided tests.

An alternative approach, designed to retain good power regardless of whether the predictor is weakly or strongly persistent, is considered in Elliott *et al.* (2015) [hereafter EMW]. EMW note that as the local-to-unity parameter $c \to \infty$, the predictive regression essentially reduces to a standard time-series regression with a weakly dependent regressor. Consequently, standard likelihood-based inference, in particular a test comparing the regression t-ratio with standard normal critical values, is an appropriate methodology. They therefore propose a hybrid test which switches between a test based on a weighted average (local asymptotic) power criterion valid when c is "small" but reverts to a standard time-series test when c is "large". In practice the choice of switching function is necessarily arbitrary; EMW propose a switching rule based on an estimate of c. The weighted average power criterion test adopted by EMW is computationally involved, and the test is also based on the assumption that the predictor cannot be locally explosive (i.e. negative values of c are not allowed), an assumption not required for the tests of CY, Kostakis *et al.* (2015) or Breitung and Demetrescu (2015).

In this paper we explore further how one can develop an approach to predictive regression testing which retains both good size properties and strong power profiles regardless of the degree of persistence of the predictor. Our approach is focused on easy to implement tests using regression t-ratios. In the near-unit root case, we base our proposed testing strategy on the use of two t-statistics: the first is the standard t-ratio test discussed above, the second is one where the predictor has been demeaned using the quasi-GLS demeaning method of Elliott *et al.* (1996), rather than OLS demeaning as with the standard t-ratio. The limiting null distributions of these statistics depend on both the endogeneity correlation parameter and the local-to-unity parameter characterising the predictor. We therefore propose a feasible method for obtaining asymptotically conservative critical values and provide response surfaces for practical use. An analysis of the asymptotic local power functions of the resulting conservative tests shows that in the empirically most relevant case where a significant negative correlation exists between returns and the predictor's innovations, the test for positive predictability based on quasi-GLS demeaning is significantly more powerful than that based on OLS demeaning. This relationship reverses when testing for negative predictability. Consequently, when testing for positive predictability, our recommended procedure in the near-unit root environment is to use the conservative standard t-ratio when the estimated endogeneity correlation is either positive or "small" and negative, but to use the conservative test based on the quasi-GLS t-ratio otherwise. Further, in common with EMW, if the data suggest the predictor is weakly persistent, we propose switching into the standard t-ratio test with reference to standard normal critical values. However, in contrast to EMW, we do not base our switching function on an (inconsistent) estimate of c, but rather on the familiar augmented Dickey-Fuller normalised bias coefficient unit root test, with MBIC lag selection as developed in Ng and Perron (2001). Our approach has the advantage of not needing to exclude the possibility of locally explosive predictors, and we show that our recommended procedure delivers effective finite sample size control and attractive power profiles across a wide range of correlation parameters and degrees of predictor persistence.

The remainder of the paper is organised as follows. Section 2 introduces the predictive regression model which we will consider in this paper together with the assumptions we place on this data generating process [DGP]. In section 3 we present the details of our hybrid switching-based test procedure and establish its asymptotic properties. Here we also outline our method for obtaining asymptotic critical values. In section 4 we investigate the finite sample size and power properties of our proposed hybrid test, comparing with the leading tests in the literature. These results suggest that the newly proposed hybrid test performs well and compares very favourably with extant tests, including its most obvious comparator test from EMW, offering simple yet highly effective methods for predictability testing. Section 5 contains a short empirical example using monthly U.S. stock returns data. Section 6 concludes. An on-line supplementary appendix contains a proof of Theorem 1 and additional material relating to the numerical simulation studies in sections 3.2 and 4.

2 The Predictive Regression Model

Let y_t denote the (excess) stock return in period t and x_{t-1} denote a scalar variable observed at time $t-1$ which is considered to be a putative predictor for y_t. Following Kostakis *et al.* (2015) and Jansson and Moreira (2006), among others, the predictive regression model we consider is specified as

$$y_t = \alpha_y + \beta x_{t-1} + \epsilon_{yt}, \qquad t = 2, ..., T \tag{1}$$

where x_t is an observed process, specified according to

$$x_t = \alpha_x + s_t, \qquad t = 1, ..., T \tag{2}$$

$$s_t = \phi s_{t-1} + \psi(L)\epsilon_{xt}, \qquad t = 2, ..., T \tag{3}$$

where $\psi(L) := 1 + \sum_{j=1}^{\infty} \psi_j L^j$ satisfying $\psi(1) \neq 0$ and $\sum_{j=1}^{\infty} j|\psi_j| = \bar{\psi} < \infty$, and where it is assumed that s_1 is a mean zero $O_p(1)$ random variable. The innovations $\boldsymbol{\epsilon}_t := (\epsilon_{xt}, \epsilon_{yt})'$ are assumed to form a (bivariate) martingale difference sequence with respect to the natural filtration $\boldsymbol{\mathcal{F}}_t = \sigma(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_{t-1}...)$, with covariance matrix $E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'|\boldsymbol{\mathcal{F}}_{t-1}) = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$, and where $\sup_t E[\|\boldsymbol{\epsilon}_t\|^{2+\kappa}] < \infty$ for some $\kappa > 0$, $\|\cdot\|$ denoting the Euclidean norm. We define the correlation between the innovations to be $\rho_{xy} := \sigma_{xy}/\sigma_x\sigma_y$.

Our interest in this paper centres on developing tests of the null hypothesis that y_t is not predictable by x_{t-1}, i.e. $H_0 : \beta = 0$ in (1). The alternative hypothesis is that y_t is predictable by x_{t-1}, in which case $\beta \neq 0$. Moreover these tests need to allow the shocks driving the predictor, ϵ_{xt} in (3), to be correlated with the unpredictable component of stock returns, ϵ_{yt} in (1), as occurs when $\rho_{xy} \neq 0$. As discussed in the Introduction it is important for practical purposes that the tests we develop are efficacious without knowledge of whether the predictor variable x_t

in (1) is weakly or strongly persistent. Formalizing, we therefore allow ϕ in (3) to satisfy one of the following two assumptions:

Assumption S. Strongly persistent predictor: The autoregressive parameter ϕ in (3) is local-to-unity with $\phi := 1 - cT^{-1}$ where c is a fixed constant.

Assumption W. Weakly persistent predictor: The autoregressive parameter ϕ in (3) is fixed and bounded away from unity, $|\phi| < 1$.

Remark 1. Many commonly used predictors are strongly persistent, exhibiting sums of sample autoregressive coefficients which are close to or only slightly smaller than unity. Near-integrated asymptotics have been found to provide better approximations for the behaviour of test statistics in such circumstances; see, *inter alia*, Elliott and Stock (1994). However, not all (putative) predictors are strongly persistent and a large part of the literature works with models which take x_t to be generated from a stable autoregressive process; see, for example, Amihud and Hurvich (2004). We therefore allow for either of these possibilities to hold for x_t. □

Remark 2. Assumption S also allows for the case where $c > 0$ such that x_t is locally explosive. While some predictive regression tests in the literature, including the tests proposed in EMW and Lewellen (2004), impose the condition that $c \leq 0$ (equivalently, $\phi \leq 1$), CY, p.54, provide a discussion on why it might not be sensible to restrict c to be non-positive in practice. Moreover, in their empirical analysis CY find that many of the predictors they consider, most notably the dividend-price ratio, have confidence intervals for ϕ that include values greater than 1. This may well be a result of local explosivity in the price series, as is well documented in the literature on financial bubbles; see, among others, Phillips *et al.* (2011, 2015). □

Remark 3. The conditions placed on the errors above essentially coincide with Assumption INNOV(i) of Kostakis *et al.* (2015, p.1512) and impose conditional homoskedasticity on $\boldsymbol{\epsilon}_t$. This is done to simplify our presentation, but it would be possible to allow for conditional heteroskedasticity of the form considered in Assumption A.1 of CY without altering the large sample results which follow under Assumption W. Under Assumption S, as in Kostakis *et al.* (2015, p.1516), an assumption of the form given in their INNOV(ii), *op. cit.*, p.1512, would be needed and the predictive regression t-ratios we discuss in the section 3 would need to be implemented using White standard errors rather than OLS standard errors. □

3 Regression-Based Predictability Tests

The simplest possible regression-based test for $H_0 : \beta = 0$ is based on the t-ratio associated with the OLS estimate of β from (1). Defining $\hat{\alpha}_y := (T-1)^{-1} \sum_{t=2}^{T} y_t$ and $\hat{\alpha}_x := (T-1)^{-1} \sum_{t=2}^{T} x_{t-1}$

this is identical to the t-statistic associated with the OLS estimate of β in the regression

$$(y_t - \hat{\alpha}_y) = \beta(x_{t-1} - \hat{\alpha}_x) + v_t, \qquad t = 2, ..., T \tag{4}$$

which is therefore defined as

$$\mathcal{T} := \frac{\sum_{t=2}^{T}(x_{t-1} - \hat{\alpha}_x)(y_t - \hat{\alpha}_y)}{\sqrt{\hat{\sigma}_v^2 \sum_{t=2}^{T}(x_{t-1} - \hat{\alpha}_x)^2}} \tag{5}$$

where $\hat{\sigma}_v^2$ is the usual OLS residual variance estimate from (4).

The representation in (4) serves to make clear two very important aspects of the basic statistic $\mathcal{T}$. First, $\mathcal{T}$ is based on separate OLS demeaning of both y_t and x_{t-1}. Under weak persistence of x_t, $\hat{\alpha}_x$ is a consistent estimator of α_x, while $\hat{\alpha}_x = O_p(T^{1/2})$ under strong persistence. Second, $\mathcal{T}$ is based on estimation that takes no account of the endogeneity present between the predictor and the regression error in (1) and, as we will see in Theorem 1, has a limiting null distribution that, under Assumption S, depends on both c and on ρ_{xy} when $c \neq 0$. In contrast, under Assumption W, where x_t is weakly persistent, $\mathcal{T}$ has a standard normal limiting null distribution and is asymptotically optimal under Gaussianity; see Jansson and Moreira (2006,p.704).

Under strong persistence, the literature to date has largely focused on the endogeneity issue. As discussed in the Introduction, analogous tests to $\mathcal{T}$ based on instrumental variable estimation of (1) have been considered in, among others, Kostakis *et al.* (2015) and Breitung and Demetrescu (2015). Other approaches which are more powerful when x_t is strongly persistent, including Lewellen (2004) and CY, fall within the general control variable approach outlined in Elliott (2011). Here (1) is augmented by an additional regressor used as a proxy for the current period innovation driving the predictor, ϵ_{xt}.[1] As discussed in Jansson and Moreira (2006,p.691), such procedures are asymptotically biased (so power can fall below the nominal level for alternatives sufficiently close to the null) as a result. The most popular example of this approach is CY's Q test. This is based around the infeasible t-statistic on β when $(x_t - \phi x_{t-1})$ is added as a regressor to (1). CY develop a feasible version of this test, using the approach of Cavanagh *et al.* (1995), based on a Bonferroni confidence interval for β formed from the sequence of such statistics across ϕ and a confidence interval for ϕ (equivalently c) formed from the well-known quasi-GLS demeaned augmented Dickey-Fuller [ADF] unit root statistic of Elliott *et al.* (1996). While Jansson and Moreira (2006) develop asymptotically uniformly most powerful tests which are asymptotically unbiased, their simulation results show that the Q test of CY has higher power in finite samples for most alternatives.

As noted above, the standard t-ratio $\mathcal{T}$ is based on OLS demeaning of both y_t and x_{t-1}.

[1] A proxy is needed because ϵ_{xt} is unobservable as both α_x, the unconditional mean of x_t in (2), and the autoregressive parameter, ϕ, in (3) are unknown. These parameters cannot be estimated at a sufficiently fast rate such that a proxy based on an estimate of ϵ_{xt} delivers (under Gaussianity) an asymptotically efficient test with a standard normal limiting null distribution, as would obtain if α_x and ϕ were known.

It is, however, well known in the literature that where a series is strongly persistent it can be advantageous to quasi-GLS demean it, as proposed in Elliott *et al* (1996), rather than use OLS demeaning. Indeed, CY adopt the quasi-GLS demeaned ADF statistic in constructing the Bonferroni-type confidence interval for c that forms the basis of their predictability test, arguing that they do so because of the superior power local power properties of the quasi-GLS demeaned ADF test relative to the standard OLS demeaned ADF test. In particular, Elliott *et al.* (1996,p.814) comment that "... where a deterministic mean or trend is present, power can be improved considerably over the standard Dickey-Fuller test by modifying the method employed to estimate the parameters characterizing the deterministic term." Elliott *et al.* (1996) develop a class of feasible near-efficient unit root tests. But the asymptotic local power functions of these tests are essentially indistinguishable from the asymptotic local power function of an ad hoc quasi-GLS demeaned regression-based ADF test despite this test not being based on any formal optimality criterion; see, in particular Figures 2 and 3 of Elliott *et al.* (1996,pp.823-4). Similarly, the MZ_{α}^{GLS} test of Ng and Perron (2001), although again not based on any formal optimality criterion, also has an asymptotic local power function that is indistinguishable from the near-efficient tests of Elliott *et al.* (1996) and superior power than the corresponding MZ_{α} test based on OLS demeaning considered in Stock (1999) and Perron and Ng (1996).

One may therefore conclude that it is largely the quasi-GLS method of demeaning that brings about this power advantage over the standard OLS demeaned unit root tests. Indeed, as Elliott *et al.* (1996,p.823-24) argue "Since the difficulties with the standard tests are associated with inefficient estimates of the trend parameters, it is reasonable to expect that modified estimates could improve their performance." It therefore seems worth investigating whether the same applies in the current situation. Consequently, rather than focusing on predictive tests in the strongly dependent case that are driven by a formal asymptotic optimality property, we will explore whether, and if so in what settings, using quasi-GLS demeaning of the persistent predictor can deliver tests with good power. Indeed, as will be shown in Theorem 1 below, under strong persistence the limiting distribution of $\mathcal{T}$ features a component which is a weighted combination of two distributions, the first of which is the local alternative limit distribution of the OLS-demeaned Dickey-Fuller statistic and the second is standard normal. The Dickey-Fuller component dominates the standard normal component when the degree of endogeneity $|\rho_{xy}|$ is large. Where the degree of endogeneity is small the reverse holds and so here we might not necessarily expect to see any gains from using a test based on quasi-GLS demeaning the predictor. As we will see in section 3.2, an exploration of the asymptotic local power functions of (asymptotically) conservative implementations (needed to account for the dependence on c and ρ_{xy} under the null) of the tests shows that quasi-GLS demeaning of the persistent predictor can indeed deliver power gains relative to $\mathcal{T}$ for moderate to large ρ_{xy} in the strongly persistent case.

To define the t-ratio from the predictive regression where the predictor regressor is quasi-GLS demeaned, we first need to define the quasi-GLS estimate of α_x. This is obtained from the

OLS regression of $(x_1, x_2 - \bar{\phi}x_1, ..., x_T - \bar{\phi}x_{T-1})$ on $(1, 1-\bar{\phi}, ..., 1-\bar{\phi})$ where $\bar{\phi} := 1 - \bar{c}/T$ with $\bar{c} = 7$; see Elliott *et al.* (1996) for further details. We denote this estimator $\tilde{\alpha}_x$. Under strong persistence $\tilde{\alpha}_x = O_p(1)$ and, hence, it is not divergent, unlike its OLS counterpart $\hat{\alpha}_x$. Based on the quasi-GLS demeaned predictor, we can define the corresponding t-statistic associated with the OLS estimate of β in regression

$$(y_t - \hat{\alpha}_y) = \beta(x_{t-1} - \tilde{\alpha}_x) + v_t \tag{6}$$

which is, hence, defined as

$$\mathcal{T}' := \frac{\sum_{t=2}^{T}(x_{t-1} - \tilde{\alpha}_x)(y_t - \hat{\alpha}_y)}{\sqrt{\tilde{\sigma}_v^2 \sum_{t=2}^{T}(x_{t-1} - \tilde{\alpha}_x)^2}} \tag{7}$$

where $\tilde{\sigma}_v^2$ is the OLS residual variance estimate from (6). Notice that we retain OLS demeaning of y_t because, under the null, $y_t = \alpha_y + \epsilon_{yt}$ is not strongly persistent and so GLS demeaning would not be appropriate.

3.1 Asymptotic Distributions of $\mathcal{T}$ and $\mathcal{T}'$

In this subsection we consider the asymptotic behaviour of the $\mathcal{T}$ and $\mathcal{T}'$ statistics. Predictive regressions for stock returns typically exhibit a small R^2 and low signal-to-noise ratios (see, *inter alia*, Campbell, 2008, and Phillips, 2015) so that departures from the null, should predictability be present, are likely to be small. Consequently, we will establish the large sample behaviour of the tests under local alternatives such that the slope parameter β in (1) is local-to-zero. The localization rate (or Pitman drift) will need to be such that β is specified to lie in a neighbourhood of zero which shrinks with the sample size, T. The appropriate Pitman drift is dictated by whether x_t is strongly or weakly persistent. Where x_t is strongly persistent, such that Assumption S holds, the appropriate local alternative is given by $H_{1,S} : \beta = gT^{-1}\sqrt{\sigma_y^2/\omega_x^2}$, where g is a finite constant and where $\omega_x^2 := \sigma_x^2\psi(1)^2$ is the long run variance of x_t. For weakly persistent x_t, such that Assumption W holds, the appropriate local alternative is given by $H_{1,W} : \beta = gT^{-1/2}\sqrt{\sigma_y^2/\psi_x^2}$, where ψ_x^2 is the short run variance of x_t, and g is again a finite constant. The different localization rates reflect the fact that near-integration implies a much stronger signal from the predictor x_{t-1}.

In Theorem 1 we now report the asymptotic distributions of the $\mathcal{T}$ and $\mathcal{T}'$ statistics under both the null and local alternatives for the case where x_t is strongly persistent. In Theorem 2, the proof of which is entirely straightforward and is therefore omitted, we subsequently present the corresponding limit for $\mathcal{T}$ for the case where x_t is weakly persistent.

Theorem 1. *Let y_t and x_t be generated according to the model in (1) -(3) under the conditions stated in Section 2 and let Assumption S hold. Let the regression t-statistics $\mathcal{T}$ and $\mathcal{T}'$ be as* *defined in* (5) *and* (7)*, respectively. Then, as* $T \to \infty$*, under* $H_{1,S}$*:*

$$(i)\ \mathcal{T} \quad \Rightarrow \quad g\sqrt{\int_0^1 \bar{W}_{1c}(r)^2 dr} + \frac{\int_0^1 \bar{W}_{1c}(r) d\left\{\rho_{xy} W_1(r) + \sqrt{1-\rho_{xy}^2} W_2(r)\right\}}{\sqrt{\int_0^1 \bar{W}_{1c}(r)^2 dr}} =: \mathcal{S}(g, \rho_{xy}, c)$$

$$(ii)\ \mathcal{T}' \quad \Rightarrow \quad g\frac{\int_0^1 \bar{W}_{1c}(r)^2 dr}{\sqrt{\int_0^1 W_{1c}(r)^2 dr}} + \frac{\int_0^1 \bar{W}_{1c}(r) d\left\{\rho_{xy} W_1(r) + \sqrt{1-\rho_{xy}^2} W_2(r)\right\}}{\sqrt{\int_0^1 W_{1c}(r)^2 dr}} =: \mathcal{S}'(g, \rho_{xy}, c)$$

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

Remark 4. Theorem 1 highlights that for both the statistics considered the offset seen in their limiting distributions under the local alternative $H_{1,S}$, and hence their asymptotic local power, is a function of the drift parameter g and a different statistic-specific stochastic offset term. Under the null hypothesis, H_0, the asymptotic distributions of both statistics are non-standard and depend on ρ_{xy} and c. When $\rho_{xy} = 0$, $\mathcal{T}$ has a $N(0,1)$ limiting distribution under H_0.

Remark 5. The limiting null distribution of $\mathcal{T}$ is seen from the representation in Theorem 1 to be a weighted average of two components, the first $\int_0^1 \bar{W}_{1c}(r) dW_1(r) / (\int_0^1 \bar{W}_{1c}(r)^2 dr)^{1/2}$ is the local alternative limit of the OLS-demeaned Dickey-Fuller statistic, while the second $\int_0^1 \bar{W}_{1c}(r) dW_2(r) / (\int_0^1 \bar{W}_{1c}(r)^2 dr)^{1/2}$ is, as noted in Remark 4, a standard $N(0,1)$ distribution. The former dominates this weighted average when $|\rho_{xy}|$ is large, while the latter dominates when $|\rho_{xy}|$ is small. Consequently, and as discussed earlier, where the degree of endogeneity is small we would not anticipate the possibility of any gains from quasi-GLS, rather than OLS, demeaning of the predictor, but where significant endogeneity is present this possibility exists. We will explore this further in section 3.2 by comparing the asymptotic local power properties of (asymptotically) size controlled tests based on $\mathcal{T}$ and $\mathcal{T}'$ under strong persistence, across a range of values of the endogeneity parameter ρ_{xy}.

Theorem 2. *Let* y_t *and* x_t *be generated according to the model in (1) -(3) under the conditions stated in Section 2 and let Assumption W hold. Then, as* $T \to \infty$*, under* $H_{1,W}$*,* $\mathcal{T} \Rightarrow N(g, 1)$*.*

Remark 6. The result in Theorem 2 demonstrates that, under Assumption W, $\mathcal{T}$ has a standard normal limiting null distribution. Notice that, unlike under Assumption S, the local power offset under $H_{1,W}$ is deterministic and equals the drift parameter, g. Indeed, as noted in Jansson and Moreira (2006,p.704), under Assumption W the test based on $\mathcal{T}$ is asymptotically optimal under Gaussianity. We do not present the corresponding limiting distribution for $\mathcal{T}'$ under Assumption W because it can be shown to depend on the distribution of s_1. The hybrid testing scheme, denoted $\mathcal{T}_{hyb}$, that we will subsequently develop in section 3.3, is designed such that it never selects $\mathcal{T}'$ in large samples under Assumption W and, hence, we will not need the limiting distribution of $\mathcal{T}'$ to establish the limiting distribution of $\mathcal{T}_{hyb}$ under Assumption W.

3.2 Asymptotic Size and Local Power Comparisons of $\mathcal{T}$ and $\mathcal{T}'$ under Strong Persistence

Under strong persistence, we can use the limiting representations given in Theorem 1 to compare the asymptotic sizes and asymptotic local powers of tests based on the $\mathcal{T}$ and $\mathcal{T}'$ statistics for a range of values of the relevant nuisance parameters on which these depend, ρ_{xy} and c. For a given value of ρ_{xy}, the main issue is that the asymptotic critical values of $\mathcal{T}$ and $\mathcal{T}'$ depend on c, which is unknown, but, unlike ρ_{xy}, is not consistently estimable. To make asymptotic size and, subsequently, asymptotic power comparisons meaningful, we adopt a scheme for simulating critical values that will, by design, deliver asymptotically conservative tests. We will illustrate this in the context of a one-sided upper-tailed test for the alternative of $\beta > 0$, but the same approach can be used in an obvious way for lower-tailed and two-tailed tests.

The steps to obtaining asymptotically conservative critical values for tests based on $\mathcal{T}$ and $\mathcal{T}'$ are as follows:

1. For a given value of ρ_{xy}, simulate the null distributions $\mathcal{S}(0, \rho_{xy}, c)$ and $\mathcal{S}'(0, \rho_{xy}, c)$ for different c across an interval $c \in [c_{\min}, c_{\max}]$.

2. At each value of c, compute the respective λ-level upper-tail critical values, $cv_\lambda(\rho_{xy}, c)$ and $cv'_\lambda(\rho_{xy}, c)$ say.

3. Set the λ-level critical values for $\mathcal{T}$ and $\mathcal{T}'$ equal to $cv_\lambda(\rho_{xy}) := \max_{c\in[c_{\min},c_{\max}]} cv_\lambda(\rho_{xy}, c)$ and $cv'_\lambda(\rho_{xy}) := \max_{c\in[c_{\min},c_{\max}]} cv'_\lambda(\rho_{xy}, c)$.

Using $cv_\lambda(\rho_{xy})$ and $cv'_\lambda(\rho_{xy})$ will yield correct λ-level sized tests based on $\mathcal{T}$ and $\mathcal{T}'$ in the case where $c = \arg\max_{c\in[c_{\min},c_{\max}]} cv_\lambda(\rho_{xy}, c)$ and $c = \arg\max_{c\in[c_{\min},c_{\max}]} cv'_\lambda(\rho_{xy}, c)$, respectively, and give conservatively sized tests for all other values of c. We simulated critical values in this manner for a significance level $\lambda = 0.05$, approximating the Brownian motion processes in the limiting functionals using $IIDN(0, 1)$ random variates, with the integrals approximated by normalised sums of 1,000 steps based on 10,000 Monte Carlo replications. This was carried out for $c_{\min} = -5$ and $c_{\max} = 50$ on the grid $c \in \{c_{\min}, c_{\min} + 1, ..., c_{\max} - 1, c_{\max}\}$, thereby covering the locally explosive, unit root and local to unit root cases. For ρ_{xy} we consider the grid $\rho_{xy} \in \{-0.950, -0.925, -0.900, ..., 0.900\}$. We will refer to the two tests where $\mathcal{T}$ and $\mathcal{T}'$ are compared with their asymptotically conservative critical values as $\mathcal{T}_{con}$ and $\mathcal{T}'_{con}$, respectively.

Figure 1 graphs the asymptotic sizes (i.e. $g = 0$) and asymptotic powers for $g = 10$ for $\mathcal{T}_{con}$ and $\mathcal{T}'_{con}$. Sizes are plotted across $c \in \{-5, -4, ...49, 50\}$ and powers across $c \in \{0, 1, ...49, 50\}$ (we do not include negative values of c here to prevent them dominating the local power plots). Panels (a)-(b), (c)-(d), (e)-(f) and (g)-(h) report results for the representative values $\rho_{xy} = -0.9$, -0.5, 0 and 0.5, respectively, while Figure S1 in the Supplementary Appendix reports results for the larger set of values $\rho_{xy} = \{-0.9, -0.7, -0.5, -0.3, -0.1, 0, 0.1, 0.3, 0.5, 0.7, 0.9\}$.

For future reference, we also show the asymptotic size of $\mathcal{T}$ evaluated at its 0.05-level critical value appropriate under weak persistence or when $\rho_{xy} = 0$, i.e. -1.645; this is denoted $\mathcal{T}_N$.

Consider first the case where $\rho_{xy} = -0.9$. Regarding the asymptotically conservative tests, $\mathcal{T}_{con}$ and $\mathcal{T}'_{con}$, we observe that $\mathcal{T}_{con}$ maximises its asymptotic size (i.e., has asymptotic size of 0.05) for c just below 0. Importantly, it is also generally very undersized for positive c. In contrast, $\mathcal{T}'_{con}$, which maximises its asymptotic size at $c = 0$, has a very flat size profile across c, never dropping much below 0.05. The pattern of asymptotic size behaviour in $\mathcal{T}_N$ essentially magnifies the pattern observed with $\mathcal{T}_{con}$, but with very bad oversize for small c. In terms of local power, between $\mathcal{T}_{con}$ and $\mathcal{T}'_{con}$, it is clear that $\mathcal{T}'_{con}$ offers substantially more power unless c is very small, with $\mathcal{T}_{con}$ suffering due to its undersize outside of the small c range. The local power plot for $\mathcal{T}_N$ is not meaningful here because of its severe oversize.

When $\rho_{xy} = -0.5$, the main feature we observe is that $\mathcal{T}_{con}$ is now less undersized for positive c compared to when $\rho_{xy} = -0.9$ (and consequently $\mathcal{T}_N$ is less oversized), while the size behaviour of $\mathcal{T}'_{con}$ is little changed. On comparing the powers, we see that $\mathcal{T}_{con}$ remains somewhat less powerful than $\mathcal{T}'_{con}$ unless c is small, although the deficit is somewhat reduced.

Results for $\rho_{xy} = 0$ show that $\mathcal{T}_{con}$ has size independent of c and coincides exactly with that of $\mathcal{T}_N$ at 0.05, however $\mathcal{T}'_{con}$ tends towards undersize unless c is large. In terms of power, this behaviour translates into $\mathcal{T}_{con}$ (i.e. $\mathcal{T}_N$) being more powerful than $\mathcal{T}'_{con}$ unless c is large, where the powers of $\mathcal{T}_{con}$ and $\mathcal{T}'_{con}$ are very close to each other.

Finally, when $\rho_{xy} = 0.5$, both $\mathcal{T}_{con}$ and $\mathcal{T}'_{con}$ tend to be undersized for smaller c, with $\mathcal{T}'_{con}$ slightly more so. However, the power gains of $\mathcal{T}_{con}$ over $\mathcal{T}'_{con}$ remain evident for the smaller values of c. Also, we see that $\mathcal{T}_N$ always has slightly lower size than $\mathcal{T}_{con}$, and so as a consequence its powers are always slightly lower.

One obvious feature is the substantial asymmetry of the size and power profiles of $\mathcal{T}_{con}$ (and $\mathcal{T}_N$) and $\mathcal{T}'_{con}$ between $\rho_{xy} = 0.5$ and $\rho_{xy} = -0.5$. For $\rho_{xy} = -0.5$, $\mathcal{T}'_{con}$ clearly possesses a better power profile than $\mathcal{T}_{con}$ while the opposite is true for $\rho_{xy} = 0.5$. This pattern of $\mathcal{T}'_{con}$ displaying better overall power properties for substantially negative ρ_{xy} and $\mathcal{T}_{con}$ outperforming $\mathcal{T}'_{con}$ for positive ρ_{xy} extends to the expanded set of ρ_{xy} results reported Figure S1 of the Supplementary Appendix, and on the basis of these it appears that $\mathcal{T}'_{con}$ has arguably the better power properties whenever $\rho_{xy} < -0.1$ and that $\mathcal{T}_{con}$ is better otherwise. While these results and conclusions are drawn for the single point $g = 10$ under the alternative, similar patterns of relative power performance arise for other values of g, as seen in Figure S2 of the Supplementary Appendix where results for $g = 5$ and $g = 20$ are reported, reinforcing the general result that $\mathcal{T}'_{con}$ has better power when $\rho_{xy} < -0.1$ and $\mathcal{T}_{con}$ better otherwise. In unreported simulations, we also found that similar relative power patterns are obtained when using the 0.10 and 0.01 significance levels.

While we have focused the foregoing analysis on upper-tail testing against the alternative $\beta > 0$, lower-tail testing of the alternative $\beta < 0$ can be carried out in an analogous fashion.

To test $\beta < 0$, for $\mathcal{T}_{con}$ and $\mathcal{T}'_{con}$ we simply replace $cv_\lambda(\rho_{xy})$ and $cv'_\lambda(\rho_{xy})$ with $-cv_\lambda(-\rho_{xy})$ and $-cv'_\lambda(-\rho_{xy})$, respectively. Consequently, an identical pattern of asymptotic sizes and powers obtains for lower-tailed tests but with ρ_{xy} replaced with $-\rho_{xy}$. That is, for lower-tailed tests, $\mathcal{T}'_{con}$ has better overall power when $\rho_{xy} > 0.1$ with $\mathcal{T}_{con}$ superior otherwise. Two-sided λ-level tests can be conducted by taking the union of rejections of the lower-tail and upper-tail tests, each conducted at the $(\lambda/2)$-level.

3.3 A Hybrid Testing Procedure

We now propose a hybrid testing procedure that is designed to capitalize on the power optimality of the standard t-test $\mathcal{T}_N$ under weak persistence, and the relative local power advantages of $\mathcal{T}_{con}$ and $\mathcal{T}'_{con}$ under strong persistence for different values of ρ_{xy}. Specifically, we consider an approach that, for upper-tail testing (lower-tail testing), (i) uses $\mathcal{T}_{con}$ under strong persistence if $\rho_{xy} > -0.1$ ($\rho_{xy} < 0.1$), (ii) uses $\mathcal{T}'_{con}$ under strong persistence if $\rho_{xy} < -0.1$ ($\rho_{xy} > 0.1$), (iii) uses $\mathcal{T}_N$ under weak persistence (for both upper- and lower-tail testing). Below we detail how to operationalise such an approach for practical implementation. We require the use of two switching mechanisms. Part (iii) involves a switching approach similar to that of EMW, whereby the standard test $\mathcal{T}_N$ is selected when evidence of a weakly persistent predictor is present. In the absence of such evidence, a secondary switching mechanism is needed to determine whether $\mathcal{T}_{con}$ or $\mathcal{T}'_{con}$ should be applied, this time on the basis of a consistent estimate of ρ_{xy}.

In the first switching mechanism, we determine whether the predictor variable is strongly or weakly persistent on the basis of a standard unit root test. For the unit root test we use the augmented Dickey-Fuller normalised bias coefficient unit root test

$$ADF_\pi := \frac{T\hat{\pi}}{1 - \sum_{i=1}^{p} \hat{\gamma}_{i,p}}$$

where $\hat{\pi}$ and $\hat{\gamma}_i$, $i = 1, ..., p$ are obtained from the estimated OLS ADF regression equation

$$\Delta x_t = \hat{\mu} + \hat{\pi} x_{t-1} + \sum_{i=1}^{p} \hat{\gamma}_i \Delta x_{t-i} + \hat{\epsilon}_{xt}. \tag{8}$$

The lag truncation parameter, p, in (8) needs to satisfy the standard rate condition that as $T \to \infty$, $1/p + p^3/T \to 0$. In practice, the lag length p can be selected by any suitable information criterion. In the numerical work which follows we will use the modified Bayes information criterion [MBIC] of Ng and Perron (2001) as we found this to deliver the best finite sample performance among popularly used lag selection rules. In the context of the MBIC rule, we used the modification suggested by Perron and Qu (2007), and the maximum permitted lag order p was set to $p_{\max} = \lfloor 12(T/100)^{1/4} \rfloor$ ($\lfloor . \rfloor$ denoting the integer part), as in Ng and Perron (2001).

Under Assumption S, $ADF_\pi = O_p(1)$, while under Assumption W, ADF_π diverges to minus infinity. Consequently, employing any fixed critical value for ADF_π, cv^{ADF} say, would ensure that $\mathcal{T}_N$ would be selected asymptotically under weak persistence since $\Pr(ADF_\pi < cv^{ADF}) \to 1$. However, in finite samples we found such a cut-off rule can lead to $\mathcal{T}_N$ being selected too often under strong persistence, leading to over-sizing of the resulting hybrid procedure. We therefore implement $\mathcal{T}_{hyb}$ with a sample size dependent critical value, $cv_T^{ADF} = -4T^{1/2}$, a choice motivated from extensive Monte Carlo simulation evidence for a range of values of T, ρ_{xy} and c. Under weak persistence, ADF_π diverges to infinity at a rate faster than $T^{1/2}$, hence $\mathcal{T}_N$ is selected asymptotically under weak persistence since $\Pr(ADF_\pi < cv_T^{ADF}) \to 1$.

In the second switching mechanism, which is operational whenever weak persistence is not detected, selection between $\mathcal{T}_{con}$ and $\mathcal{T}'_{con}$ is made on the basis of the ρ_{xy} estimator

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

where the $\hat{\epsilon}_{yt}$ are the OLS residuals from regressing y_t on a constant and x_{t-1}, and $\hat{\epsilon}_{xt}$ are the ADF residuals from (8). This estimator is consistent for ρ_{xy} under either Assumption S or Assumption W. In practice then, for $\mathcal{T}_{con}$ and $\mathcal{T}'_{con}$ we use the critical values $cv_\lambda(\hat{\rho}_{xy})$ and $cv'_\lambda(\hat{\rho}_{xy})$ as estimates of $cv_\lambda(\rho_{xy})$ and $cv'_\lambda(\rho_{xy})$. To automate selection of an appropriate critical value we calculated a response surface by OLS regressions of $cv_\lambda(\rho_{xy})$ and $cv'_\lambda(\rho_{xy})$ on $[1, \rho_{xy}, \rho_{xy}^2, ..., \rho_{xy}^8]$ for the grid of values $\rho_{xy} = \{-0.90, -0.85, ..., 0.9\}$ (37 data points). The response surface coefficient estimates are given in Table 1 for the usual values of λ (the R^2 from the response surface regressions exceeded 0.999 in all cases), and the response surface critical value is obtained as the fitted value from the corresponding estimated regression.

To summarise, our suggested hybrid double switching-based testing procedure, which we denote by $\mathcal{T}_{hyb}$, is defined as follows:

1. If $ADF_\pi < -4T^{1/2}$ perform $\mathcal{T}_N$ ($\mathcal{T}$ with a standard normal critical value).
2. Otherwise:
   (a) For upper-tail tests against the alternative $\beta > 0$,

   if $\hat{\rho}_{xy} > -0.1$ perform $\mathcal{T}_{con}$ ($\mathcal{T}$ with conservative critical value $cv_\lambda(\hat{\rho}_{xy})$)
   if $\hat{\rho}_{xy} < -0.1$ perform $\mathcal{T}'_{con}$ ($\mathcal{T}'$ with conservative critical value $cv'_\lambda(\hat{\rho}_{xy})$)

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

   if $\hat{\rho}_{xy} < 0.1$ perform $\mathcal{T}_{con}$ ($\mathcal{T}$ with conservative critical value $-cv_\lambda(-\hat{\rho}_{xy})$)
   if $\hat{\rho}_{xy} > 0.1$ perform $\mathcal{T}'_{con}$ ($\mathcal{T}'$ with conservative critical value $-cv'_\lambda(-\hat{\rho}_{xy})$)

In the next section we explore the efficacy of this hybrid testing approach in delivering a procedure with reliable size and attractive power, relative to existing tests in the literature, across a wide range of correlation parameters ρ_{xy} and degrees of predictor persistence.

4 Finite Sample Size and Power

We examine the finite sample size and power properties of the $\mathcal{T}_{hyb}$ procedure and compare these with the prominent tests in the predictive regression testing literature. Specifically, the tests we employ as comparators are CY's Q test; BD, the instrumental variable test of Breitung and Demetrescu (2015) using their recommended sine and fractional instruments (denoted BD); the test of Kostakis *et al.* (2015) (denoted IVX), and the test of EMW (denoted EMW). Note that we compare with the original Q test of CY, rather than a modified variant that can control size under weak persistence, because EMW find in their supplement that the modified test has lower power than the original test for moderate values of c, and is dominated by the EMW test, and also because the original Q test is the one implemented by practitioners, hence it presents a more useful point of comparison. We do not report the test of Jansson and Moreira (2006) because, as noted earlier, the Q test has higher power than this test in finite samples for most alternatives.

We generate data for a sample size $T = 200$ from the model (1)-(3) with $(\epsilon_{xt}, \epsilon_{yt})' \sim IIDN(0, I_2)$, $\psi(L) = 1$ and drawing s_1 as a standard normal variate. We set $\alpha_y = \alpha_x = 0$ as all the tests considered are invariant to these constant terms. We examine rejection frequencies for $\phi \in \{1.025, 1, 0.975, 0.95, 0.875, 0.75, 0.5, 0\}$, thereby varying x_t between an explosive and white noise process, for $\rho_{xy} = \{-0.9, -0.5, 0, 0.5, 0.9\}$, with $\beta = 0$ and $\beta > 0$ corresponding to size and power respectively. The reported results are based on 10,000 Monte Carlo replications.

We conduct a 0.05-level upper tail test for $\mathcal{T}_{hyb}$, Q and EMW; a 0.10-level two-tailed test for BD (recall that this test can only be run as a two-tailed test) and consider two variants of IVX: a 0.05-level upper tail test and a 0.10-level two-tailed test, denoting these as IVX_1 and IVX_2 respectively. The Q tests were computed using the code provided by CY.[2] To implement EMW, we adopt their switching function so that the standard test $\mathcal{T}$ based on (4) is applied if a (non-consistent) estimate of the local offset c is at least 130, while their weighted average power criterion-based test is applied otherwise, using the sample statistics and long run correlation estimator specified on p.697 of Jansson and Moreira (2006), together with the routines provided by EMW.[3] For the estimate of c we use the natural estimator from (8), $-T\hat{\pi}$, and when the standard $\mathcal{T}$ test is used in EMW, we follow EMW's approach of setting the critical value to the usual value of 1.645 for non-negative estimates of the long run correlation parameter, but to set it

[2] The CY routines are downloadable from: `http://jfe.rochester.edu/data.htm`.

[3] The EMW routines are downloadable from: `https://www.econometricsociety.org/content/supplement-nearly-optimal-tests-when-nuisance-parameter-present-under-null-hypothesis-0`.

to 1.7 for negative estimates. In calculating the IVX_1 and IVX_2 tests we implemented the finite-sample correction factor outlined in Kostakis *et al.* (2015, p.1516). Although the innovations in (3) are generated without serial correlation, we do not assume knowledge of this when running the tests (as would be the case in practical applications). For Q we set $p_{\max} = \lfloor 12(T/100)^{1/4} \rfloor$, in line with the $p_{\max}$ setting used in $\mathcal{T}_{hyb}$, while for IVX and EMW, long run variances are calculated using a Bartlett kernel with lag truncation $\lfloor T^{1/3} \rfloor$ (BD requires no serial correlation correction).

The simulation results are shown in Figures 2-6. Considering first the sizes of the tests across the different ϕ and ρ_{xy} settings, the newly proposed $\mathcal{T}_{hyb}$ test displays excellent finite sample size control across the full range of persistence and correlation parameters, with very little deviation from nominal size, apart from some undersize for positive ρ_{xy} in the more persistent cases. Of the existing competitor tests, BD and IVX_2 also demonstrate decent size behaviour, while the remaining tests can be badly size distorted. Specifically, EMW does not control size for explosive processes, e.g. size is close to one for $\rho_{xy} = -0.9$ and close to zero for $\rho_{xy} \geq 0$; also EMW displays substantial oversize for moderate and small values of ϕ when $\rho_{xy} > 0$. On the other hand, Q is severely oversized when $\phi = 0$ and also suffers severe undersize when $\phi = 1.025$ and $\rho_{xy} = 0.5$, and IVX_1 can be badly oversized for more persistent series when ρ_{xy} is negative. That EMW does not control size for explosive processes and Q does not control size for white noise processes is not surprising given that these tests are not designed to be valid in such circumstances. However, the severe size distortions displayed for these settings highlight the sensitivity of these tests to departures from the persistence assumptions under which they were derived, and the contrast with tests such as $\mathcal{T}_{hyb}$, which offer robustness to a much broader set of persistence parameters, is stark.

Turning attention to the power performance of the procedures, for $\rho_{xy} = -0.9$ (Figure 2), we find that for the explosive setting $\phi = 1.025$, the correctly sized tests $\mathcal{T}_{hyb}$ and IVX_2 have very similar power profiles, lying only a little below those of BD and IVX_1 which are modestly oversized in this case. The power of Q is very low in comparison to the other tests here, while comparison with EMW is not meaningful due to it having a size close to one. In the unit root case $\phi = 1$, EMW dominates all other tests in terms of power; it appears that exclusion of robustness to the case of explosive predictors affords the EMW test the opportunity of greater power in the unit root setting. Of the other tests, $\mathcal{T}_{hyb}$ is next best for small departures from the null while Q offers some gains over $\mathcal{T}_{hyb}$ for larger β, while both of these tests offer significant power advantages over IVX_2 and BD (IVX_1 is oversized and hence cannot be compared in terms of power). As the process becomes less persistent, the power advantages of EMW are very quickly eliminated, with $\mathcal{T}_{hyb}$ offering the best power profile (of the correctly sized tests) even for $\phi = 0.975$. For $\rho_{xy} = -0.9$, the overall picture is one of $\mathcal{T}_{hyb}$ offering the best power profile for all values of ϕ except $\phi = 1$ where EMW dominates.

When $\rho_{xy} = -0.5$ (Figure 3), similar comments apply to the explosive and unit root cases, although in the unit root case, the power gains of EMW over $\mathcal{T}_{hyb}$ are not as marked. For $\phi = 0.975$, 0.95 and 0.875, the power profiles of EMW, $\mathcal{T}_{hyb}$ and Q essentially coincide, and are the best performing tests for these degrees of persistence. When $\phi = 0.75$ and 0.5, power gains of $\mathcal{T}_{hyb}$ over EMW and Q are seen, the magnitude of which can be quite substantial in the $\phi = 0.5$ case. For the white noise setting $\phi = 0$, $\mathcal{T}_{hyb}$ and EMW again coincide while Q is badly oversized *and* has poor power. In the case of $\rho_{xy} = 0$ (Figure 4), with the exception of the explosive case (where EMW has very low size and power), the $\mathcal{T}_{hyb}$, EMW, Q and IVX_2 tests share very similar properties (EMW offers some small power gains for the most persistent settings), while IVX_1 and BD lag behind in terms of power performance. When $\rho_{xy} = 0.5$ (Figure 5), $\mathcal{T}_{hyb}$ performs best for an explosive predictor, with EMW and Q displaying considerably lower power here, while EMW offers the best power profile for $\phi = 1$, 0.975 and 0.95. However, as the persistence parameter decreases further, EMW first becomes oversized (as noted above) with $\mathcal{T}_{hyb}$ providing the best power of the correctly sized tests, until $\phi = 0$ when the power profiles of EMW and $\mathcal{T}_{hyb}$ again coincide. Finally, for $\rho_{xy} = 0.9$ (Figure 6), a generally more exaggerated picture of the $\rho_{xy} = 0.5$ results is seen, with $\mathcal{T}_{hyb}$ dominant for $\phi = 1.025$, EMW markedly best for $\phi = 1$ and 0.975, but then EMW suffering from oversize for smaller ϕ with $\mathcal{T}_{hyb}$ being the best performing test of those correctly sized. Across Figures 3-6, $\mathcal{T}_{hyb}$ and EMW arguably emerge as the tests with the best overall power profiles, with each test offering relative power advantages over the other in different settings. However, of these two procedures, $\mathcal{T}_{hyb}$ is alone in also offering reliable size control across the full range of persistence and correlation settings.

In the Supplementary Appendix, Figures S3-S20 report results for cases where additional serial correlation is permitted in the predictor series, with s_t of (2) specified as $s_t = \phi s_{t-1} + u_t$, $u_t = \delta u_{t-1} + \epsilon_{xt} - \theta \epsilon_{x,t-1}$, with simulations conducted for $\theta = \pm 0.5$ and $\delta = \pm 0.5$. We find that $\mathcal{T}_{hyb}$ retains its feature of never being subject to large upward size distortions, in contrast to Q and EMW whose sizes can vary dramatically for different combinations of ϕ, ρ_{xy}, δ and θ. For example, Q can now display substantial oversize for $\phi = 1.025$ across all values of ρ_{xy}, as well as an increased range of less persistent ϕ cases, especially when $\rho_{xy} > 0$, while the oversize seen for EMW in Figures 2-6 for small and moderate ϕ when $\rho_{xy} > 0$ can now extend to the $\phi = 0$ case and sometimes become more pronounced. When the tests are approximately correctly sized so that meaningful power comparisons can be made, the same broad patterns emerge as in Figures 2-6, albeit with some differences in the magnitudes of the relative power gains/losses.

Based on our simulation results, we conclude that $\mathcal{T}_{hyb}$ offers appealing size and power properties when compared to the leading currently available testing procedures. It would be fairly naïve to believe, *a priori*, that any one test procedure would have the best finite sample size and power properties across the full constellation of settings that we have examined, i.e. a wide spectrum of values of the persistence level in the predictive regressor and the correlation coefficient between the innovations in the model. However, $\mathcal{T}_{hyb}$ does appear to perform consistently well in terms of both size and power across these settings, never seemingly showing a substantial weakness in either dimension, something which appears to be rather less true of its extant competitors.

5 An Empirical Illustration

To illustrate the use of our proposed test in practice, we apply it, together with its competitors, to the monthly U.S. annual equity series analysed in Welch and Goyal (2008), using updated data for the period 1980:1-2017:12 ($T = 456$) which is available at `http://www.hec.unil.ch/agoyal/`. Our dependent variable, y_t, is the S&P 500 value-weighted log excess return and for x_t we consider thirteen putative predictor variables: the dividend price ratio, earnings-price ratio, dividend-payout ratio, dividend yield, default yield spread, long-term yield, default return spread, stock variance, net equity expansion, inflation rate, Treasury bill rate, term spread and the book-to-market value ratio. Detail of the construction of these predictors can be found in Welch and Goyal (2008). The test procedures are all applied with the same settings and serial correlation corrections as used in section 4 above. We conduct one-sided upper-tail tests for $\mathcal{T}_{hyb}$, EMW, Q and IVX_1 (with the exception of the stock variance predictor for which we apply lower-tail tests), and two-sided tests for BD and IVX_2; the tests are implemented at the 0.10, 0.05 and 0.01 significance levels for $\mathcal{T}_{hyb}$, BD, IVX_1 and IVX_2, and at the 0.05-level for EMW and Q (again using the code provided by EMW and CY, respectively).

The results are presented in Table 2, along with the values of $\hat{\rho}_{xy}$ and ADF_π (in this application, $cv_T^{ADF} = -4T^{1/2} = -85.4$). There are three cases where one or more of the tests reject at the 0.10-level or above: the dividend yield, default return spread and stock variance. For the dividend yield, only $\mathcal{T}_{hyb}$ and Q reject; here $\hat{\rho}_{xy}$ is close to zero so we would expect $\mathcal{T}_{hyb}$ and Q to give similar results, although interestingly none of the other tests reject, including EMW. Due to the persistence in this predictor, as evidenced by the small value of ADF_π, $\mathcal{T}_{hyb}$ is here using $\mathcal{T}_{con}$. In the case of the default return spread, all tests but BD exhibit rejections, while for the stock variance all the tests reject. The values of ADF_π for these two predictors suggests very low levels of persistence, with $\mathcal{T}_{hyb}$ using $\mathcal{T}_N$ and EMW also switching into the standard t-test. In summary, fairly limited evidence of return predictability is found across the set of predictors considered, but it is clear that $\mathcal{T}_{hyb}$ uncovers at least as much evidence for predictability as any of its comparator tests.

6 Conclusions

We have developed new and easy to implement tests for predictability based on computationally simple regression t-ratios and a switching rule based on a conventional normalised bias ADF statistic implemented with the MBIC lag selection rule of Ng and Perron (2001). In particular, together with the standard t-ratio from the OLS regression of returns on a constant and a lagged predictor, we have discussed a t-ratio from a variant of the standard predictive regression where the OLS demeaned returns are regressed on the quasi-GLS demeaned lagged predictor. Where the predictor is strongly persistent, we have proposed a feasible method for obtaining (conservative) asymptotic critical values for tests based on each of these statistics and associated response surfaces have been provided. An analysis of the asymptotic local power functions of the resulting (asymptotically) conservative tests in the case where the predictor is strongly persistent showed that these vary considerably with the endogeneity correlation parameter. We consequently suggest applying either the conservative standard t-ratio or its quasi-GLS variant, according to the magnitude of the estimated endogeneity correlation parameter. Where the predictor is weakly persistent the standard t-ratio compared to standard normal critical values is optimal under Gaussianity. We therefore propose a switching testing procedure, similar in approach to that considered in Elliott *et al.* (2015), whereby one of the two conservative tests is performed, as outlined above, unless the normalised bias ADF statistic indicates that the predictor is weakly dependent, in which case we compare the standard t-ratio with standard normal critical values. Monte Carlo simulations presented suggest that our hybrid test compares very favourably with the leading tests for predictability in the literature, offering arguably the best trade-off of in terms of overall finite sample size and power properties across a broad diversity of persistence and endogeneity settings.

We conclude with a suggestion for further research. Like the vast majority of the published tests in this literature we have considered the case of a single predictor. Some published papers have considered multiple predictors simultaneously, most notably the IV-based tests of Kostakis *et al.* (2015) and Breitung and Demetrescu (2015). Both of these, however, assume that either all of the predictors are weakly persistent, or all of the predictors are strongly persistent, thereby disallowing sets of predictors with mixed orders of persistence. The bootstrap tests of Bauer and Hamilton (2018) also allow for multiple predictors, but again make the same assumption. Amihud *et al.* (2009) also allow for multiple predictors but these must all be weakly dependent. Under the assumption of a common order of persistence, it may be possible to generalise the approach outlined in this paper to accommodate multiple predictors. Investigating this possibility and how well it works in practice compared to the other tests mentioned above is beyond the scope of the present paper but would constitute an interesting topic for further research.

References

Amihud, Y. and C. M. Hurvich (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis* 39, 813–841.

Amihud, Y., C. M. Hurvich, and Y. Wang (2009). Multiple-predictor regressions: Hypothesis testing. *Review of Financial Studies* 22, 413–434.

Bauer, M. D. and J. D. Hamilton (2018). Robust bond risk premia. *Review of Financial Studies* 31, 399–448.

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

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics* 18, 373–400.

Campbell, J. Y. (2008). Viewpoint: Estimating the equity premium. *Canadian Journal of Economics/Revue canadienne d'économique* 41, 1–21.

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

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

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

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

Elliott, G. (2011). A control function approach for testing the usefulness of trending variables in forecast models and linear regression. *Journal of Econometrics* 164, 79–91.

Elliott, G., U. K. Müller and M. W. Watson (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica* 83, 771–811.

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

Elliott, G. and J. H. Stock (1994). Inference in time series regression when the order of integration of a regressor is unknown. *Econometric Theory* 10, 672–700.

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

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

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

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

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

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

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

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics* 74, 209–235.

Ng, S. and P. Perron (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519–1554.

Perron, P. and S. Ng (1996). Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *Review of Economic Studies* 63, 435–463.

Perron, P. and Z. Qu (2007). A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. *Economics Letters* 94, 12–19.

Phillips, P. C. B. (2015). Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics* 13, 521–555.

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

Phillips, P. C. B., S.-P. Shi and J. Yu (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the SP500. *International Economic Review* 56, 1043–1078.

Phillips, P. C. B., Y. Wu and J. Yu (2011). Explosive behavior in the 1990s Nasdaq: When did the exuberance escalate asset values? *International Economic Review* 52, 201–226.

Stock, J. H. (1999). A class of tests for integration and cointegration. In R. F. Engle and H. White, eds, *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, pp.137–167. Oxford University Press.

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

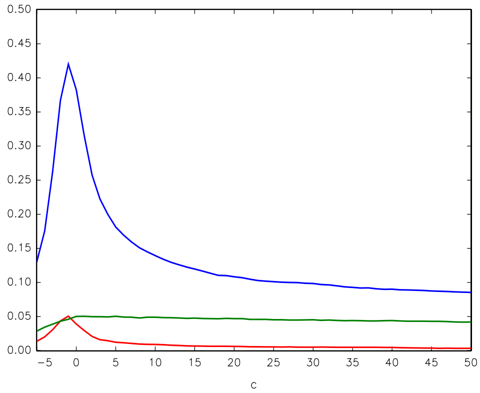
Table 1. Response surface coefficient estimates for $\mathcal{T}_{con}$ and $\mathcal{T}'_{con}$

	$cv_\lambda(\rho_{xy})$				$cv'_\lambda(\rho_{xy})$			
Regressor	$\lambda = 0.1$	$\lambda = 0.05$	$\lambda = 0.025$	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 0.05$	$\lambda = 0.025$	$\lambda = 0.01$
1	1.346	1.707	2.004	2.434	1.293	1.648	1.950	2.377
ρ_{xy}	−0.819	−0.802	−0.765	−0.726	−0.242	−0.225	−0.285	−0.382
ρ_{xy}^2	1.928	2.314	1.947	1.257	−0.055	0.323	0.200	−0.171
ρ_{xy}^3	−0.402	−0.377	−0.602	−0.736	−0.316	−0.275	−0.186	0.414
ρ_{xy}^4	−5.008	−6.970	−5.131	−2.385	0.493	−1.447	−0.559	0.209
ρ_{xy}^5	0.825	1.013	0.965	1.448	0.401	0.432	−0.005	−0.984
ρ_{xy}^6	7.040	10.279	6.692	1.972	−0.808	2.603	0.224	−0.504
ρ_{xy}^7	−0.470	−0.705	−0.350	−0.762	−0.200	−0.290	0.219	0.693
ρ_{xy}^8	−3.607	−5.417	−3.154	−0.479	0.459	−1.581	0.236	0.434

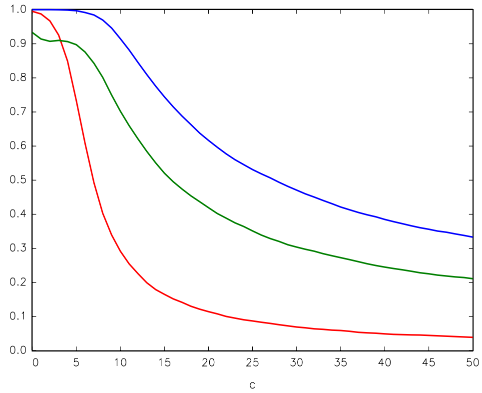
Table 2. Application to monthly U.S. S&P 500 returns, 1980:1-2017:12

Predictor	$\hat{\rho}_{xy}$	ADF_π	$\mathcal{T}_{hyb} =$	$\mathcal{T}_{hyb}$	BD	IVX_1	IVX_2	Q	EMW
Dividend payout ratio	−0.07	−39.84	$\mathcal{T}_{con}$	0.19	0.46	0.14	0.02		
Earnings-price ratio	−0.58	−16.22	$\mathcal{T}'_{con}$	0.68	0.91	1.06	1.12		
Dividend-price ratio	−0.99	−3.92	$\mathcal{T}'_{con}$	0.91	0.32	0.85	0.72		
Dividend yield	−0.04	−4.15	$\mathcal{T}_{con}$	1.80**	0.76	0.96	0.93	**	
Default yield spread	−0.13	−23.81	$\mathcal{T}'_{con}$	−0.01	0.07	−0.31	0.09		
Long-term yield	−0.13	−2.92	$\mathcal{T}'_{con}$	0.08	0.19	−0.25	0.06		
Default return spread	0.24	−280.05	$\mathcal{T}_N$	1.98**	2.37	1.80**	3.25*	**	**
Stock variance	−0.36	−126.52	$\mathcal{T}_N$	−3.19***	12.94***	−3.32***	11.00***	**	**
Net equity expansion	0.03	−17.51	$\mathcal{T}_{con}$	0.09	0.96	0.14	0.02		
Inflation rate	0.00	−44.76	$\mathcal{T}_{con}$	−0.41	0.29	−0.87	0.76		
Treasury bill rate	−0.02	−7.87	$\mathcal{T}_{con}$	0.01	0.08	−0.53	0.28		
Term spread	−0.02	−28.33	$\mathcal{T}_{con}$	0.26	0.06	0.45	0.20		
Book-to-market value ratio	−0.66	−5.77	$\mathcal{T}'_{con}$	0.38	0.02	0.23	0.05		

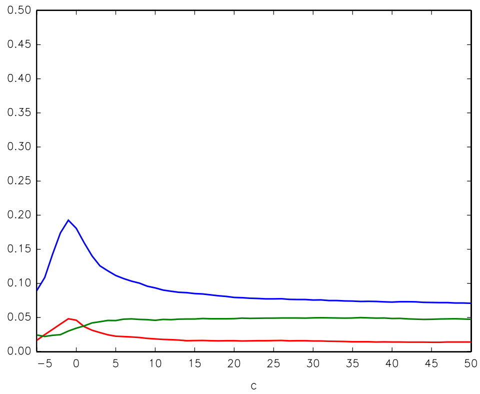
Notes: *, ** and *** denotes rejection at the 0.10-, 0.05- and 0.01-levels, respectively. The column labelled "$\mathcal{T}_{hyb} =$" states which of the constituent tests is selected in the hybrid test $\mathcal{T}_{hyb}$.



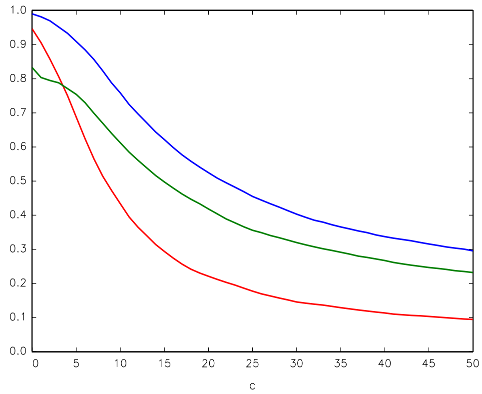
(a) $g = 0$, $\rho_{xy} = -0.9$



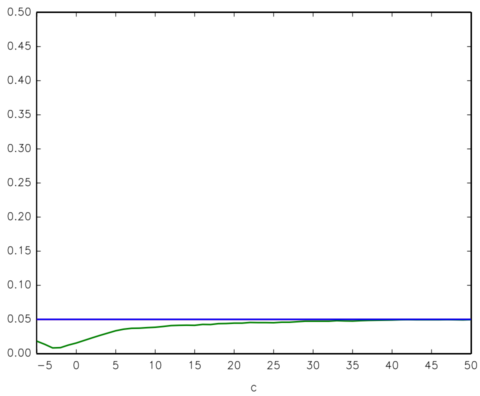
(b) $g = 10$, $\rho_{xy} = -0.9$



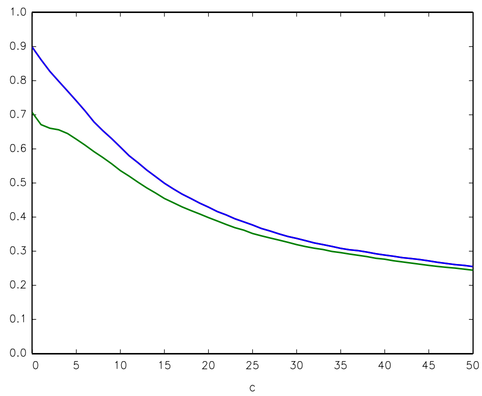
(c) $g = 0$, $\rho_{xy} = -0.5$



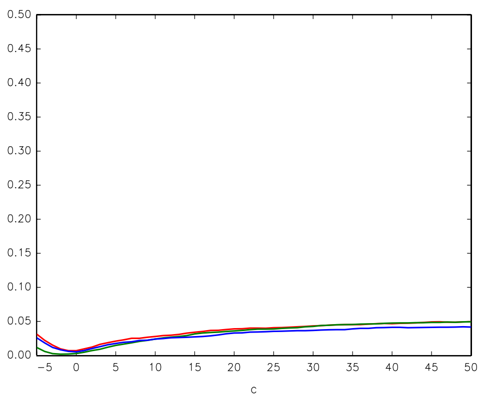
(d) $g = 10$, $\rho_{xy} = -0.5$



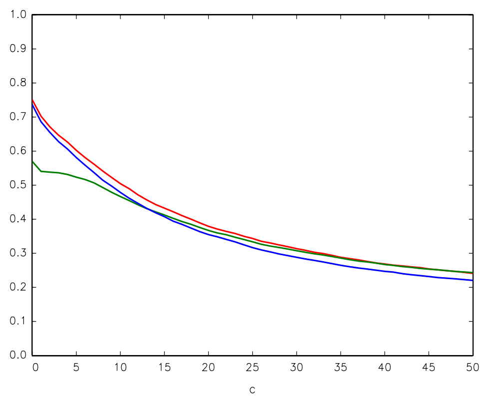
(e) $g = 0$, $\rho_{xy} = 0$



(f) $g = 10$, $\rho_{xy} = 0$

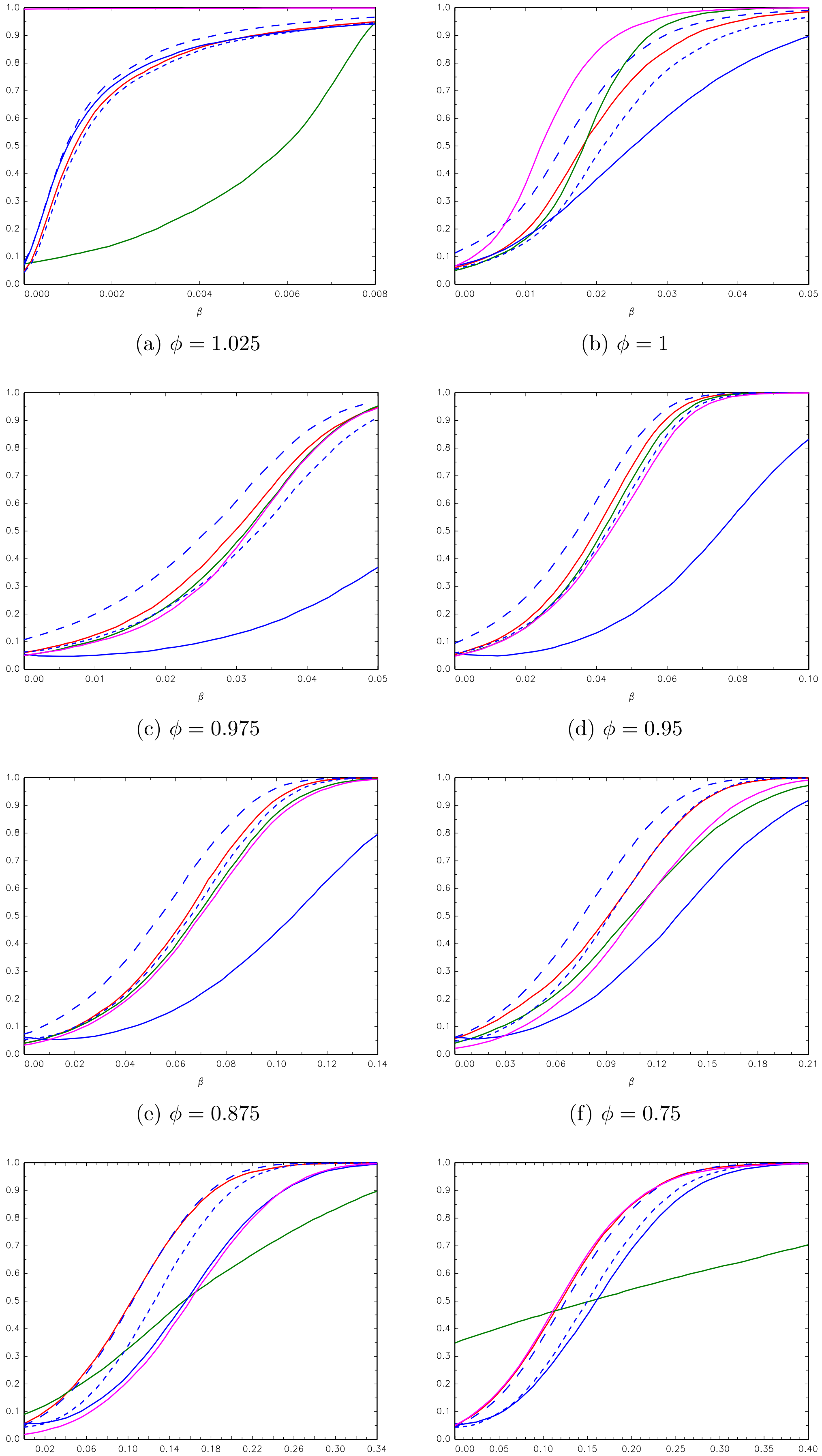


(g) $g = 0$, $\rho_{xy} = 0.5$



(h) $g = 10$, $\rho_{xy} = 0.5$

Figure 1. Asymptotic size ($g = 0$) and local power ($g = 10$) of nominal 0.05-level tests based on the model in equations (1)-(3) with innovation correlation ρ_{xy}; $\mathcal{T}_{con}$: —— , $\mathcal{T}'_{con}$: —— , $\mathcal{T}_N$: ——



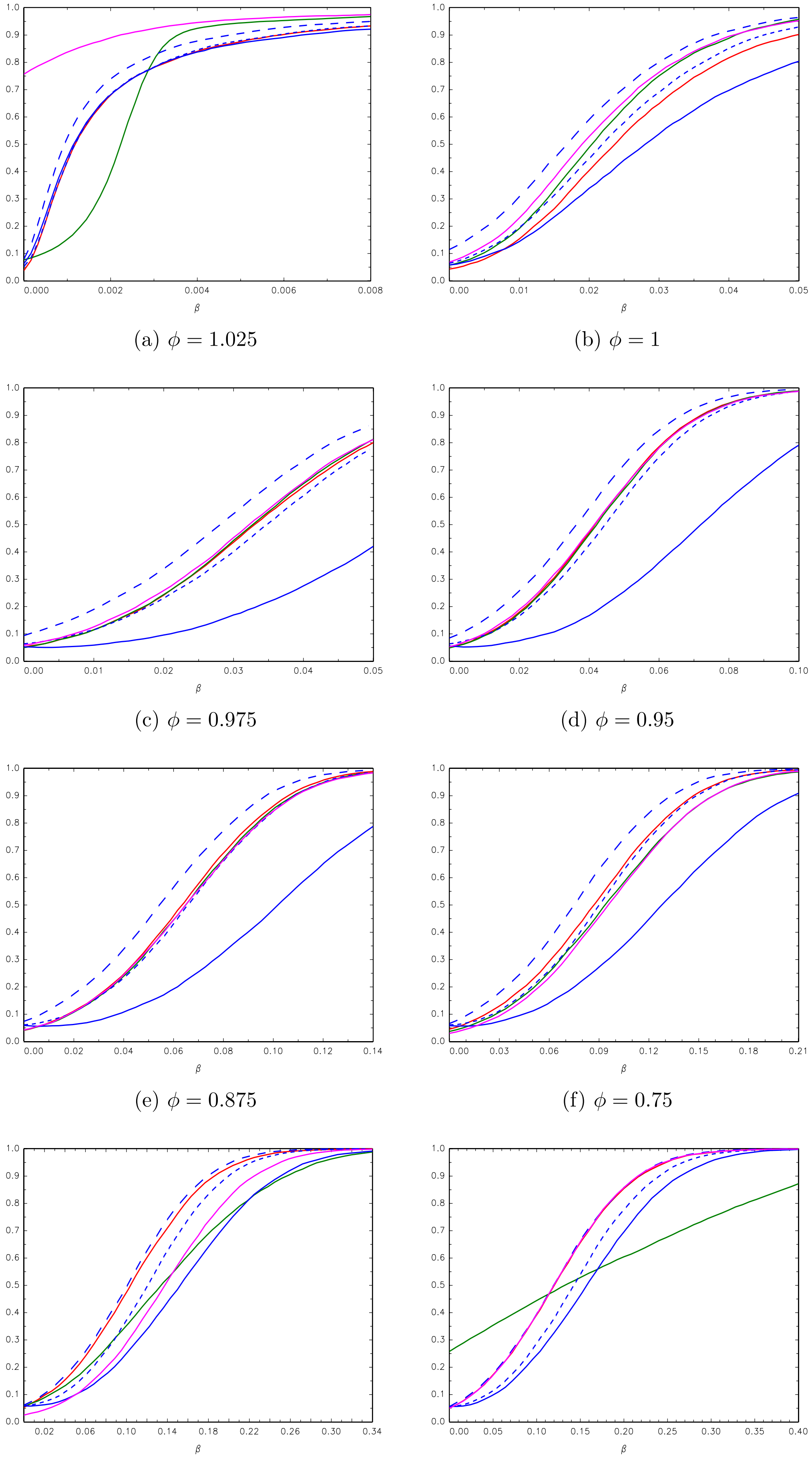
(a) $\phi = 1.025$ (b) $\phi = 1$

(c) $\phi = 0.975$ (d) $\phi = 0.95$

(e) $\phi = 0.875$ (f) $\phi = 0.75$

(g) $\phi = 0.5$ (h) $\phi = 0$

Figure 2. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.9$; $\mathcal{T}_{hyb}$: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line, EMW: magenta solid line



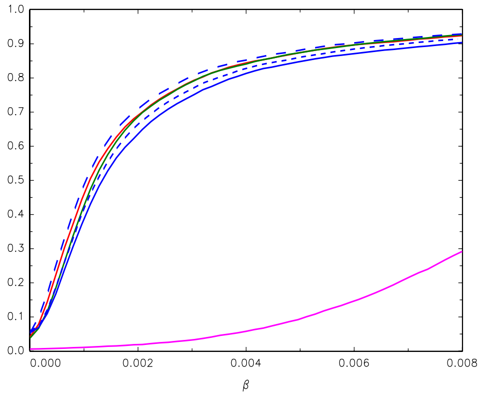
(a) $\phi = 1.025$ (b) $\phi = 1$

(c) $\phi = 0.975$ (d) $\phi = 0.95$

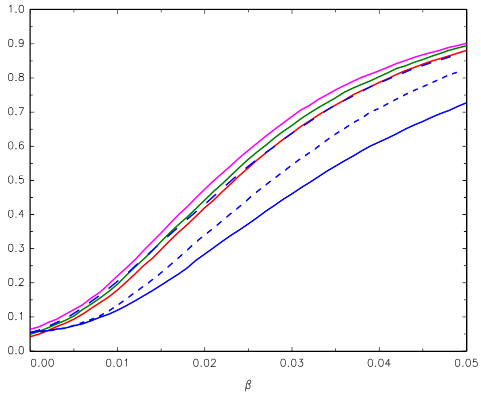
(e) $\phi = 0.875$ (f) $\phi = 0.75$

(g) $\phi = 0.5$ (h) $\phi = 0$

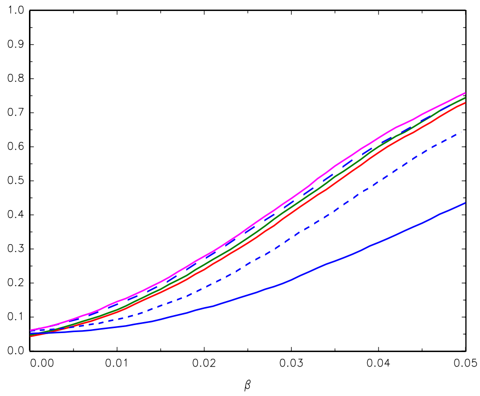
Figure 3. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.5$; $\mathcal{T}_{hyb}$: —— (red), Q: —— (green), BD: —— (blue), IVX_1: – – – (blue), IVX_2: - - - - (blue), EMW: —— (magenta)



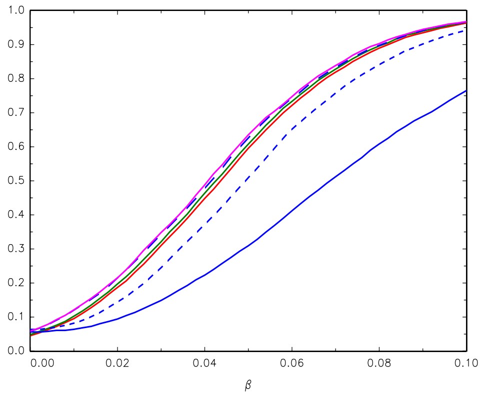
(a) $\phi = 1.025$



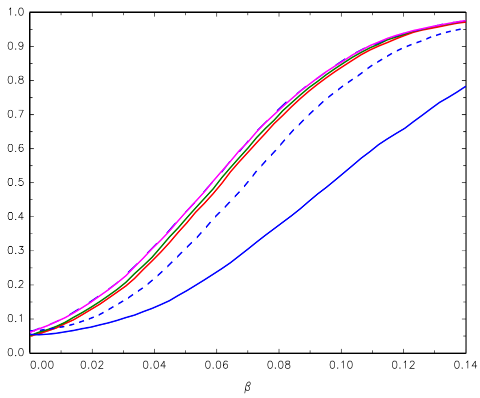
(b) $\phi = 1$



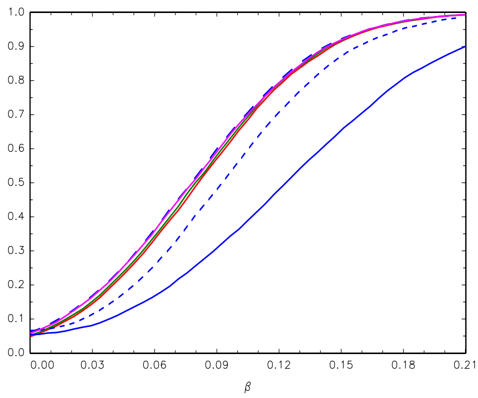
(c) $\phi = 0.975$



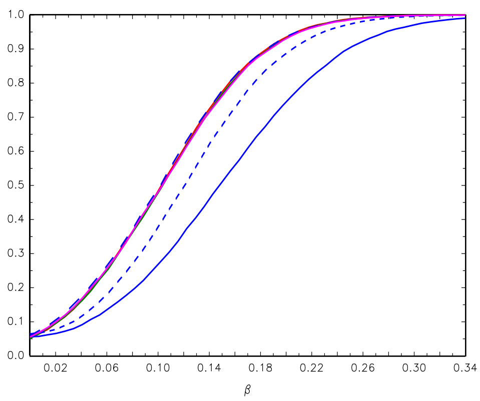
(d) $\phi = 0.95$



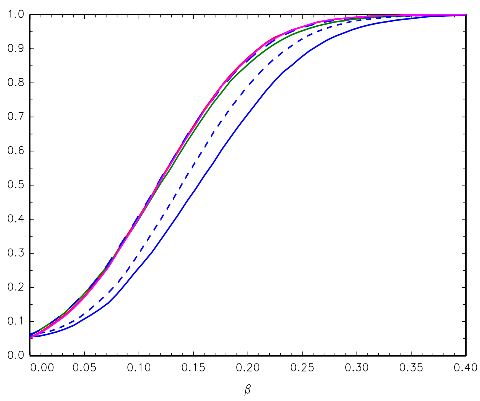
(e) $\phi = 0.875$



(f) $\phi = 0.75$

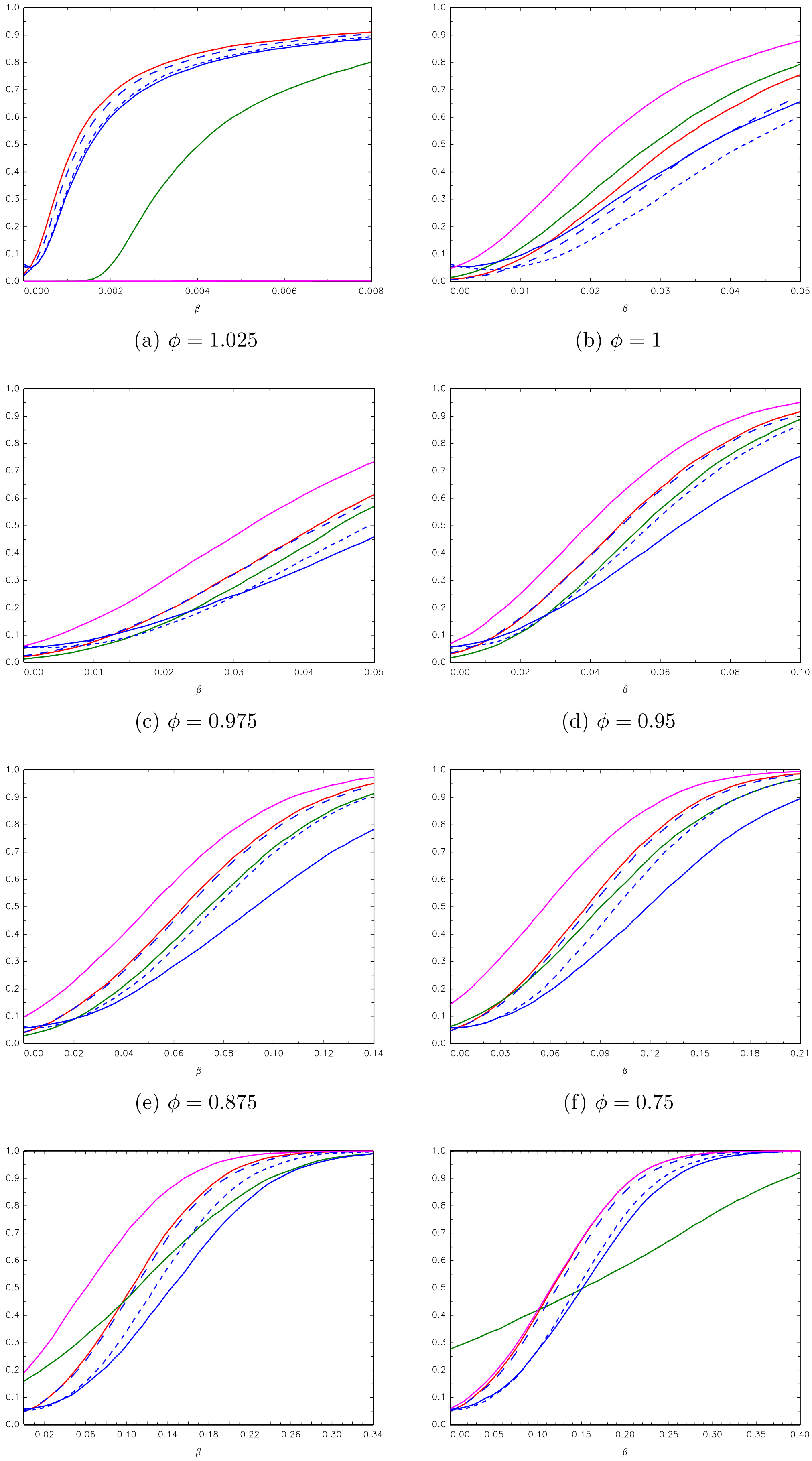


(g) $\phi = 0.5$



(h) $\phi = 0$

Figure 4. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0$; $\mathcal{T}_{hyb}$: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line, EMW: magenta solid line



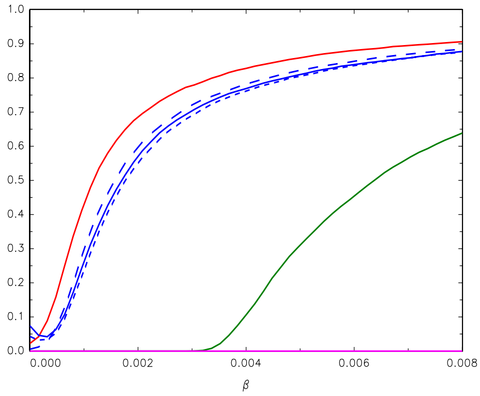
(a) $\phi = 1.025$ (b) $\phi = 1$

(c) $\phi = 0.975$ (d) $\phi = 0.95$

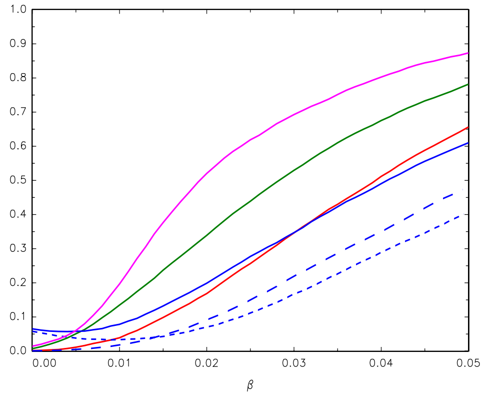
(e) $\phi = 0.875$ (f) $\phi = 0.75$

(g) $\phi = 0.5$ (h) $\phi = 0$

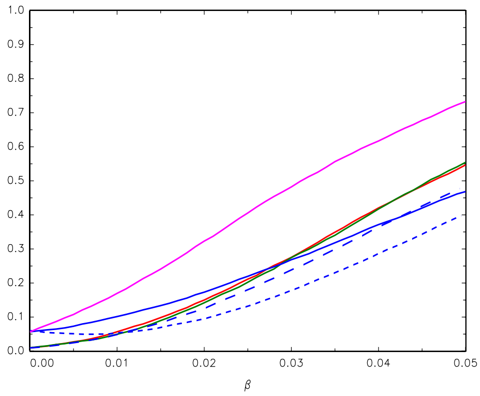
Figure 5. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.5$; $\mathcal{T}_{hyb}$: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line, EMW: magenta solid line



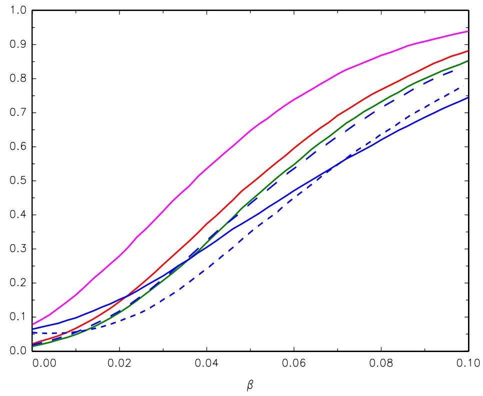
(a) $\phi = 1.025$



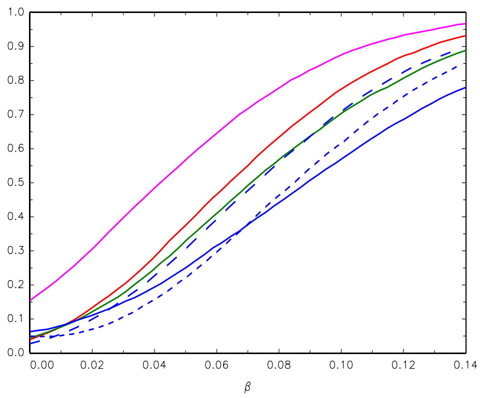
(b) $\phi = 1$



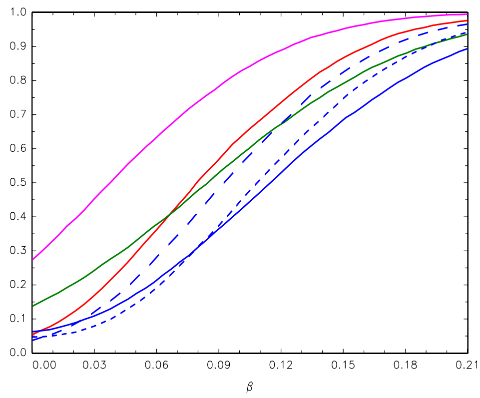
(c) $\phi = 0.975$



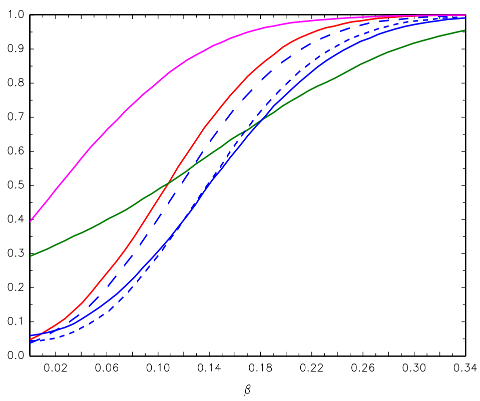
(d) $\phi = 0.95$



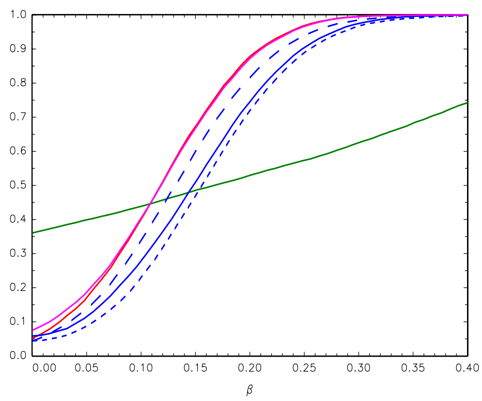
(e) $\phi = 0.875$



(f) $\phi = 0.75$



(g) $\phi = 0.5$



(h) $\phi = 0$

Figure 6. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.9$; $\mathcal{T}_{hyb}$: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line, EMW: magenta solid line

On-Line Supplementary Appendix

to

"Simple Tests for Stock Return Predictability with Good Size and Power Properties"

by

D.I. Harvey, S.J. Leybourne and A.M. Robert Taylor

Summary of Contents

This supplement to the paper "Simple Tests for Stock Return Predictability with Good Size and Power Properties " contains two sections. In section S.1 a proof of Theorem 1 is provided. Section S.2 reports additional supporting Monte Carlo results to those reported in sections 3.2 and 4. of the paper.

S.1 Proof of Theorem 1

The result given for $\mathcal{T}$ in part (i) of Theorem 1 is well known and is given in, for example, Equation (3.11) p.1140 of Cavanagh *et al.* (1995).

To show part (ii) of Theorem 1, we first note that

$$\sum_{t=2}^{T}(x_{t-1} - \tilde{\alpha}_x)(y_t - \hat{\alpha}_y) = \sum_{t=2}^{T}(x_{t-1} - \hat{\alpha}_x)(y_t - \hat{\alpha}_y)$$

since $\sum_{t=2}^{T}(y_t - \hat{\alpha}_y) = 0$. We thus find that $\mathcal{T}'$ and $\mathcal{T}$ are related according to

$$\mathcal{T}' = \sqrt{\frac{\hat{\sigma}_v^2 \sum_{t=2}^{T}(x_{t-1} - \hat{\alpha}_x)^2}{\tilde{\sigma}_v^2 \sum_{t=2}^{T}(x_{t-1} - \tilde{\alpha}_x)^2}}\mathcal{T}.$$

Then, under Assumption S,

$$\begin{aligned} T^{-2}\sum_{t=2}^{T}(x_{t-1}-\hat{\alpha}_x)^2 &\Rightarrow \omega_x^2\textstyle\int_0^1 \bar{W}_{1c}(r)^2dr, \\ T^{-2}\sum_{t=2}^{T}(x_{t-1}-\tilde{\alpha}_x)^2 &= T^{-2}\sum_{t=2}^{T}(x_{t-1}-x_1)^2+o_p(1) \\ &\Rightarrow \omega_x^2\textstyle\int_0^1 W_{1c}(r)^2dr \end{aligned}$$

and since it is straightforward to demonstrate that $\hat{\sigma}_v^2 = \sigma_y^2 + o_p(1)$ and $\tilde{\sigma}_v^2 = \sigma_y^2 + o_p(1)$, we find that

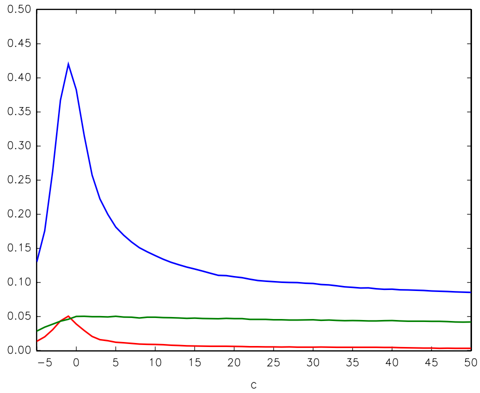
$$\begin{aligned} \mathcal{T}' &\Rightarrow \sqrt{\frac{\int_0^1 \bar{W}_{1c}(r)^2dr}{\int_0^1 W_{1c}(r)^2dr}}\mathcal{S}(g,\rho_{xy},c) \\ &= \mathcal{S}'(g,\rho_{xy},c). \end{aligned}$$

S.2 Additional Monte Carlo Results

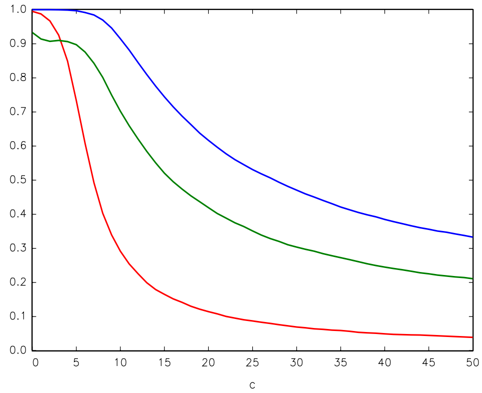
In this section we report additional supporting Monte Carlo results to those reported in sections 3.2 and 4. of the paper. Specifically:

• Figures S1 and S2 report the asymptotic size and local power of the $\mathcal{T}_{con}$, $\mathcal{T}'_{con}$ and $\mathcal{T}_N$ tests for an augmented set of values of the non-centrality parameter, g, and the endogeneity correlation parameter, ρ_{xy}, to those reported in Figure 1 in section 3.2. Specifically, for $g = 0$ and $g = 10$, Figure S.1 reports the full set of results for $\rho_{xy} = \{-0.9, -0.7, -0.5, -0.3, -0.1, 0, 0.1, 0.3, 0.5, 0.7, 0.9\}$. Figure S2 reports the corresponding set of results for $g = \{5, 20\}$ and $\rho_{xy} = \{-0.9, -0.7, -0.5, -0.3, -0.1, 0, 0.1, 0.3, 0.5, 0.7, 0.9\}$. All computational aspects are as detailed in section 3.2 in the context of the results in Figure 1.

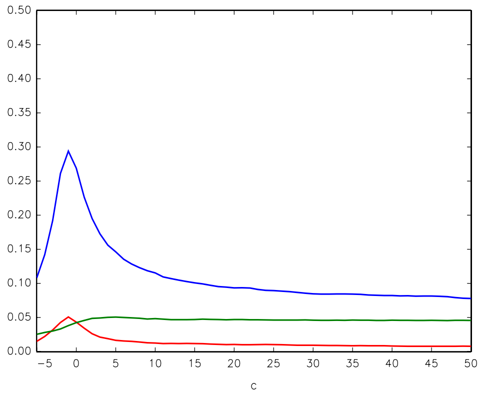
• Figures S3-S22 report finite sample power curves analogous to those reported in Figures 2-6 in section 4 in the case where additional serial correlation is permitted in the predictor series, with s_t of (2) specified as $s_t = \phi s_{t-1} + u_t$, $u_t = \delta u_{t-1} + \epsilon_{xt} - \theta\epsilon_{x,t-1}$, with simulations conducted for $\theta = \pm 0.5$ and $\delta = \pm 0.5$. All computational aspects are as detailed in section 4 in the context of the results in Figures 2-6.



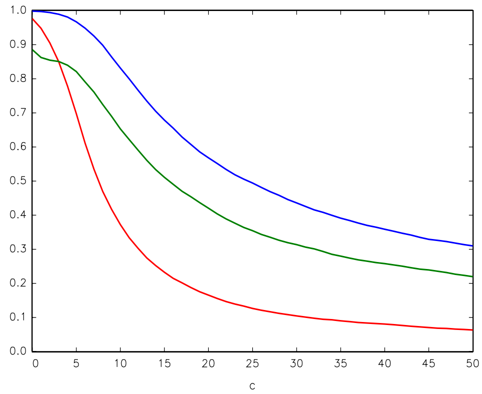
(a) $g = 0$, $\rho_{xy} = -0.9$



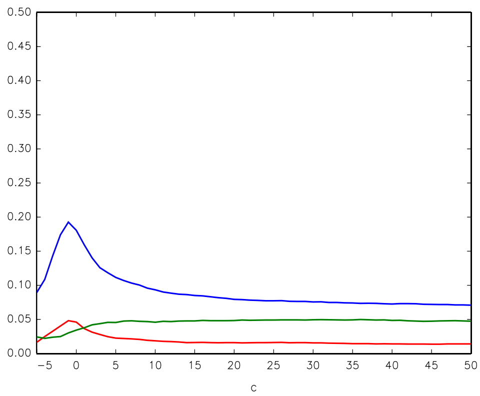
(b) $g = 10$, $\rho_{xy} = -0.9$



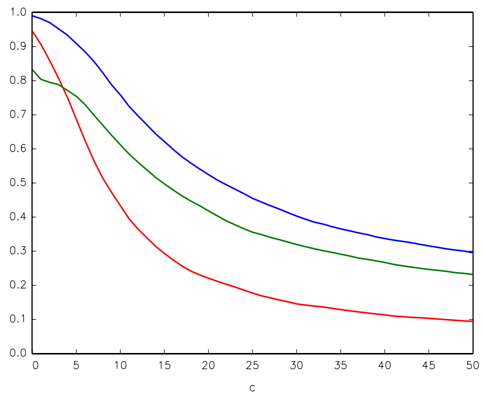
(c) $g = 0$, $\rho_{xy} = -0.7$



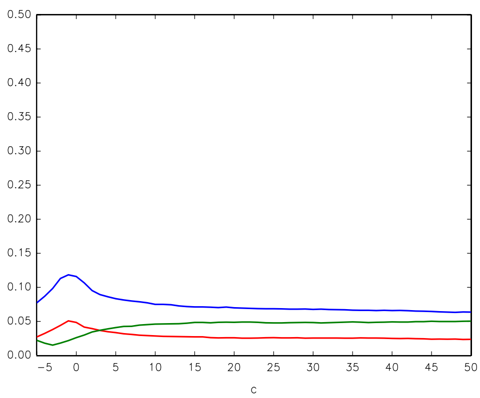
(d) $g = 10$, $\rho_{xy} = -0.7$



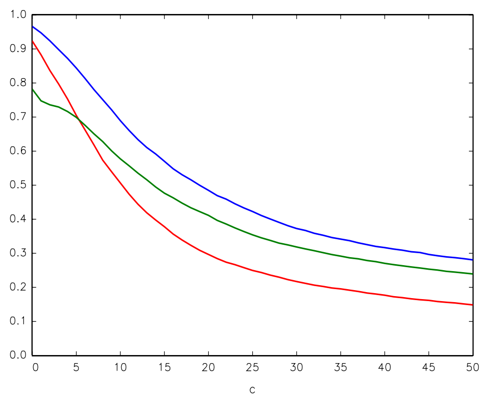
(e) $g = 0$, $\rho_{xy} = -0.5$



(f) $g = 10$, $\rho_{xy} = -0.5$

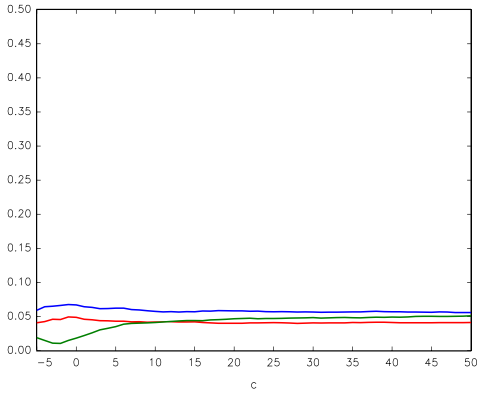


(g) $g = 0$, $\rho_{xy} = -0.3$

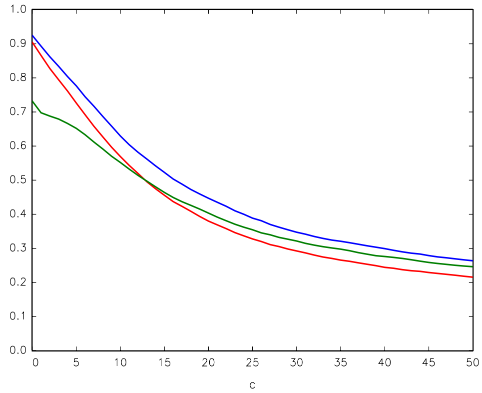


(h) $g = 10$, $\rho_{xy} = -0.3$

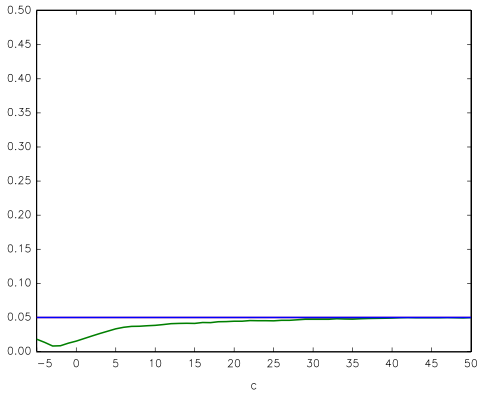
Figure S1. Asymptotic size ($g = 0$) and local power ($g = 10$) of nominal 0.05-level tests, $\mathcal{T}_{con}$: red line, $\mathcal{T}'_{con}$: green line, $\mathcal{T}_N$: blue line



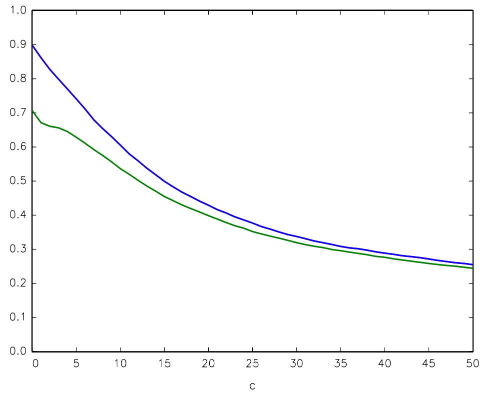
(i) $g = 0$, $\rho_{xy} = -0.1$



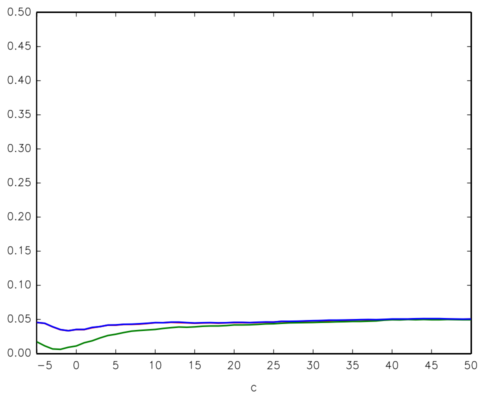
(j) $g = 10$, $\rho_{xy} = -0.1$



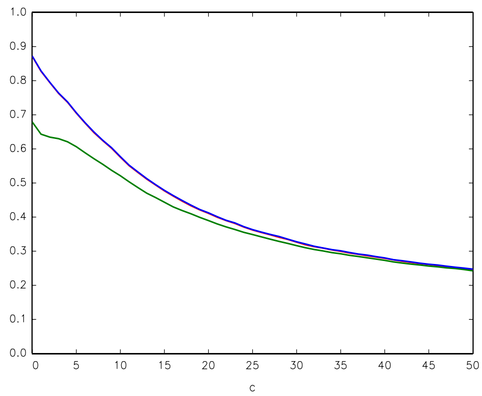
(k) $g = 0$, $\rho_{xy} = 0$



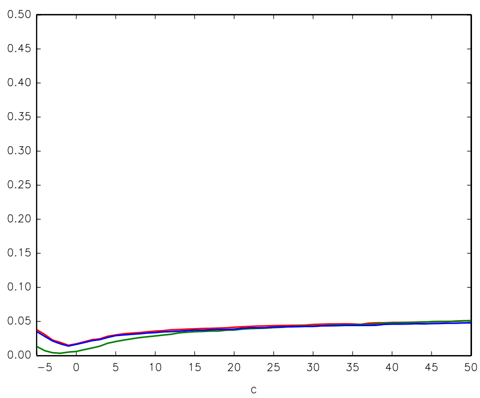
(l) $g = 10$, $\rho_{xy} = 0$



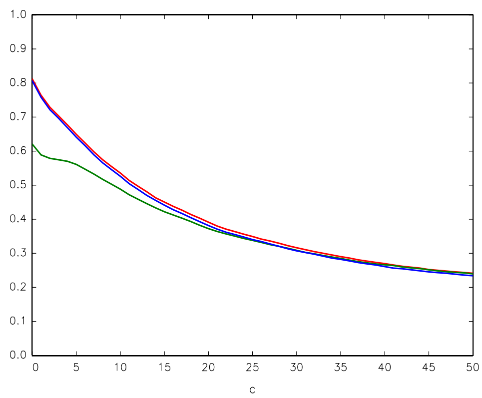
(m) $g = 0$, $\rho_{xy} = 0.1$



(n) $g = 10$, $\rho_{xy} = 0.1$

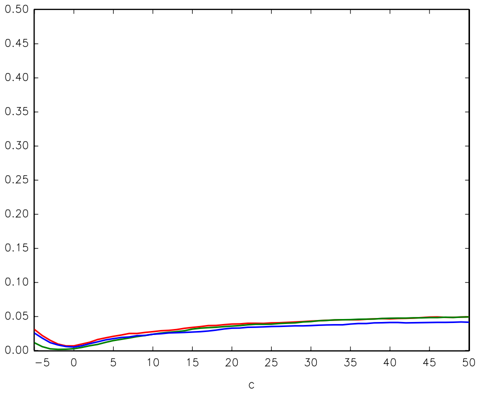


(o) $g = 0$, $\rho_{xy} = 0.3$

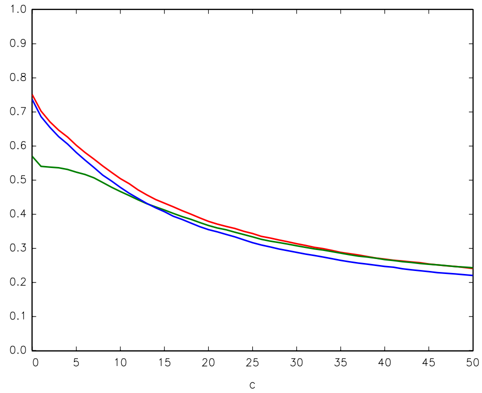


(p) $g = 10$, $\rho_{xy} = 0.3$

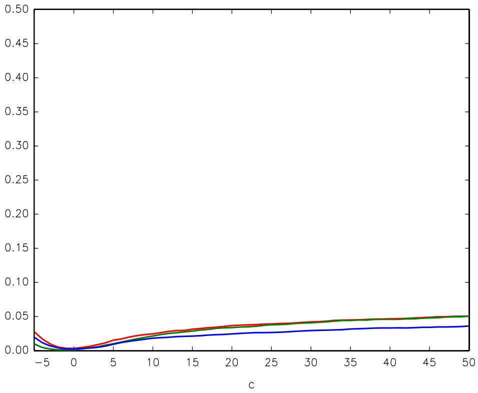
Figure S1 (continued). Asymptotic size ($g = 0$) and local power ($g = 10$) of nominal 0.05-level tests, $\mathcal{T}_{con}$: red line, $\mathcal{T}'_{con}$: green line, $\mathcal{T}_N$: blue line



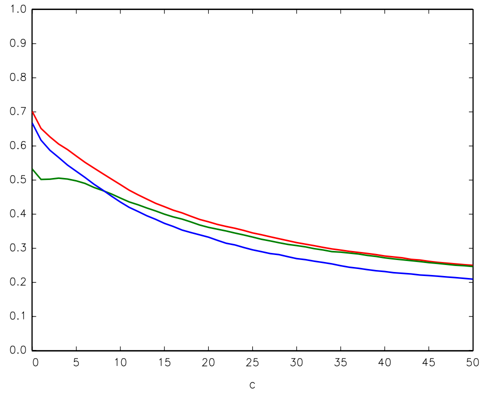
(q) $g = 0$, $\rho_{xy} = 0.5$



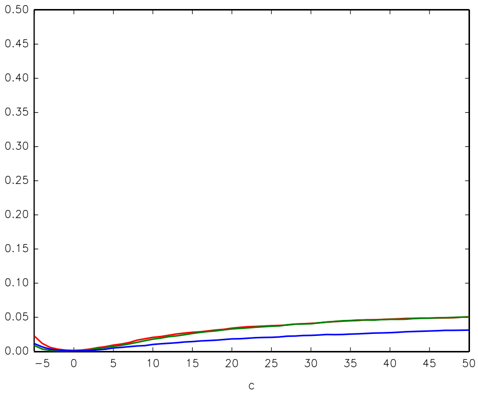
(r) $g = 10$, $\rho_{xy} = 0.5$



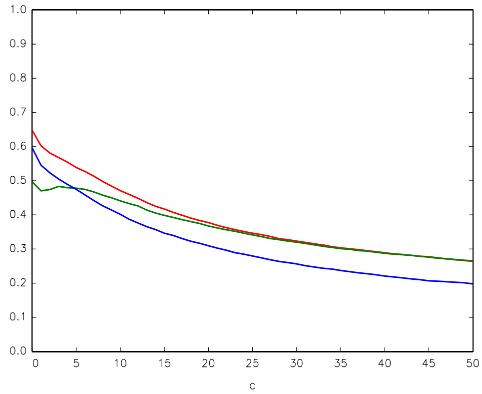
(s) $g = 0$, $\rho_{xy} = 0.7$



(t) $g = 10$, $\rho_{xy} = 0.7$

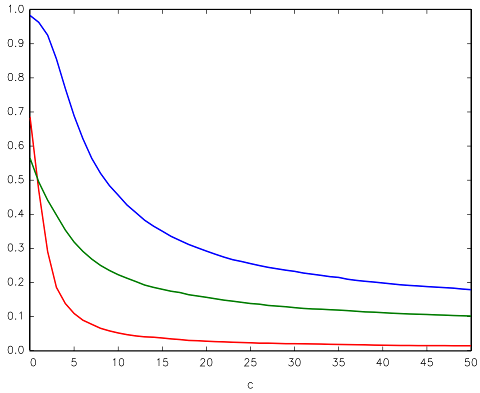


(u) $g = 0$, $\rho_{xy} = 0.9$

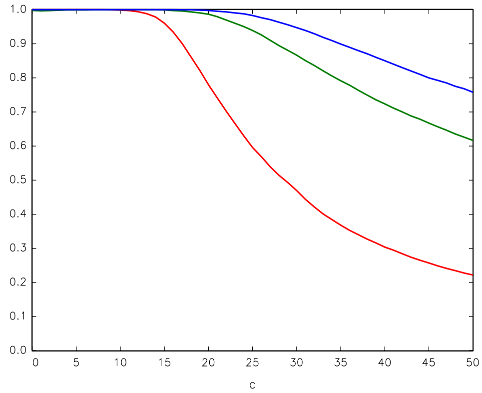


(v) $g = 10$, $\rho_{xy} = 0.9$

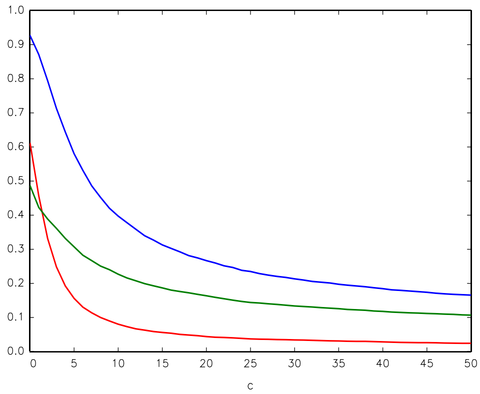
Figure S1 (continued). Asymptotic size ($g = 0$) and local power ($g = 10$) of nominal 0.05-level tests, $\mathcal{T}_{con}$: red line, $\mathcal{T}'_{con}$: green line, $\mathcal{T}_N$: blue line



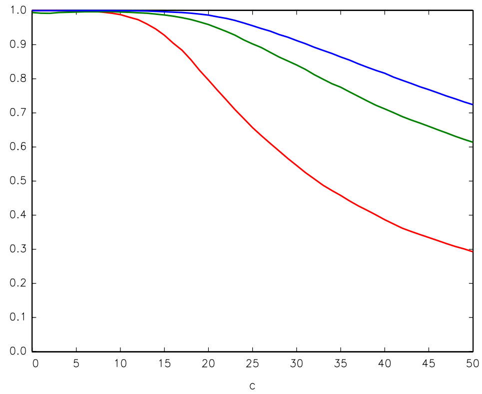
(a) $g = 5$, $\rho_{xy} = -0.9$



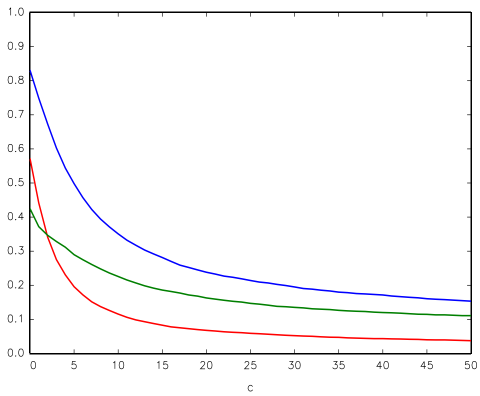
(b) $g = 20$, $\rho_{xy} = -0.9$



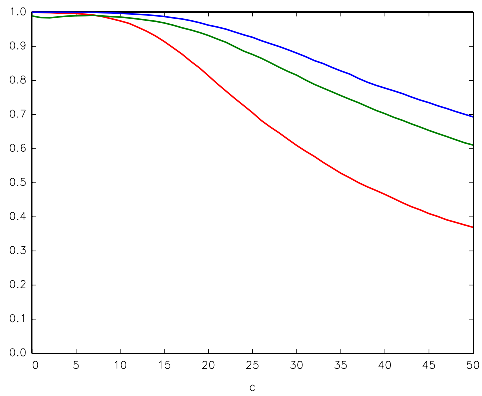
(c) $g = 5$, $\rho_{xy} = -0.7$



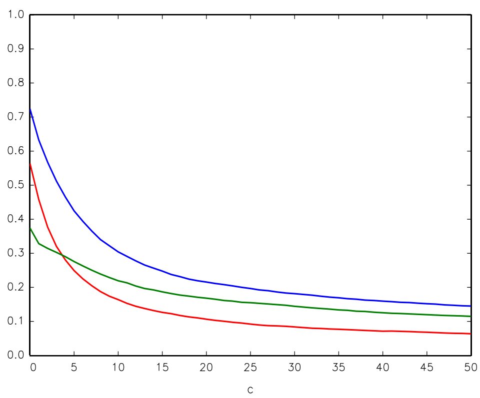
(d) $g = 20$, $\rho_{xy} = -0.7$



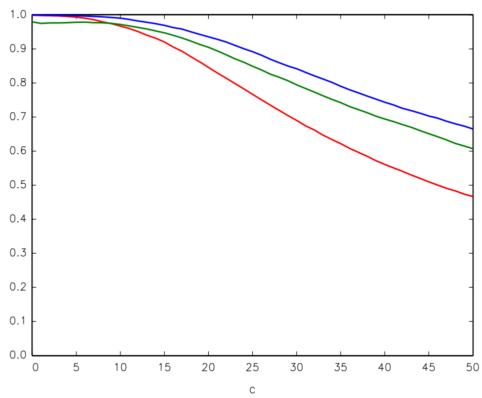
(e) $g = 5$, $\rho_{xy} = -0.5$



(f) $g = 20$, $\rho_{xy} = -0.5$

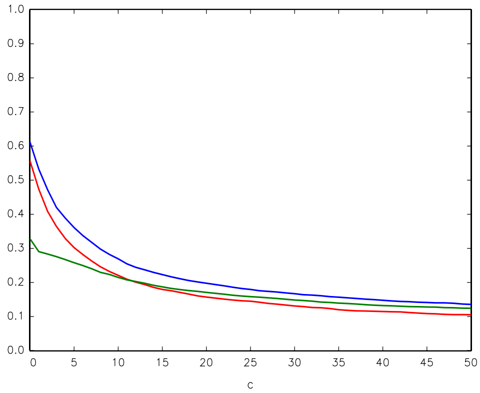


(g) $g = 5$, $\rho_{xy} = -0.3$

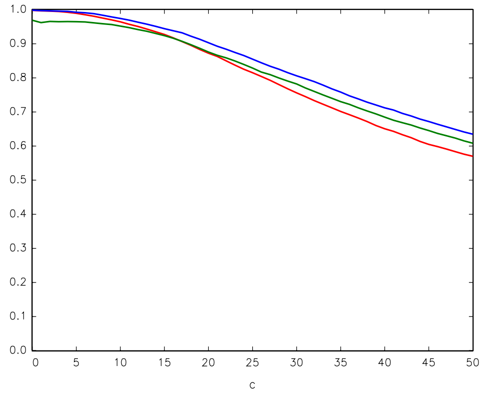


(h) $g = 20$, $\rho_{xy} = -0.3$

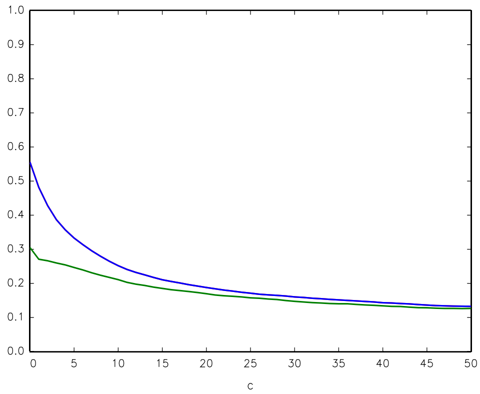
Figure S2. Asymptotic local power ($g = 5$ and $g = 20$) of nominal 0.05-level tests, $\mathcal{T}_{con}$: —— , $\mathcal{T}'_{con}$: —— , $\mathcal{T}_N$: ——



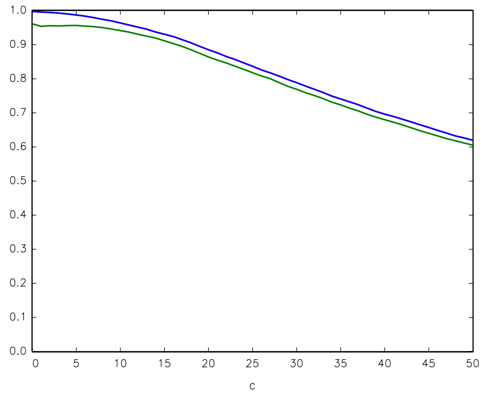
(i) $g = 5$, $\rho_{xy} = -0.1$



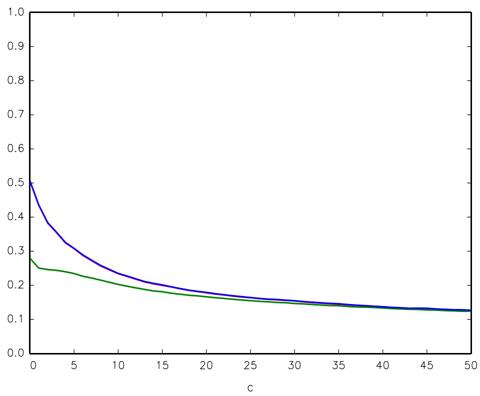
(j) $g = 20$, $\rho_{xy} = -0.1$



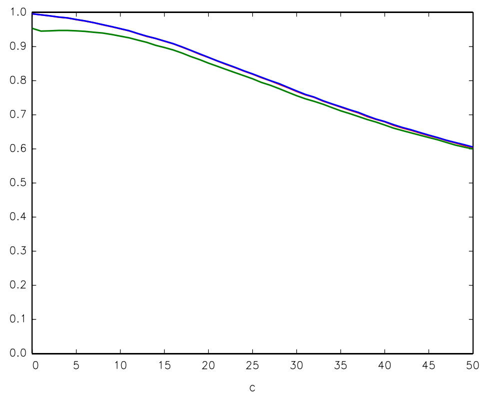
(k) $g = 5$, $\rho_{xy} = 0$



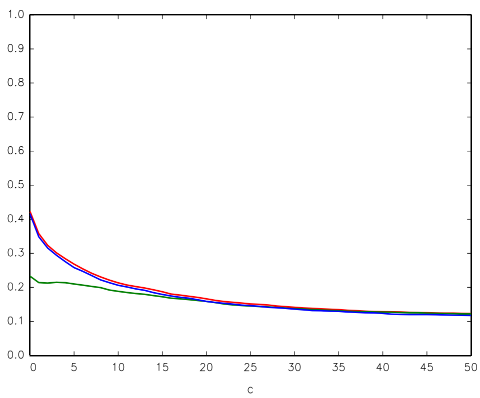
(l) $g = 20$, $\rho_{xy} = 0$



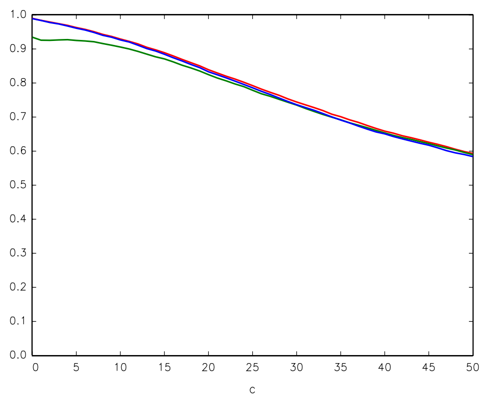
(m) $g = 5$, $\rho_{xy} = 0.1$



(n) $g = 20$, $\rho_{xy} = 0.1$

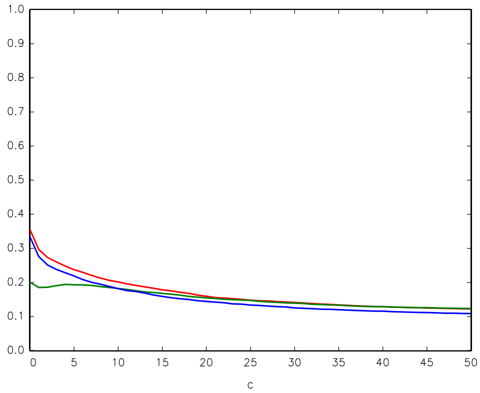


(o) $g = 5$, $\rho_{xy} = 0.3$

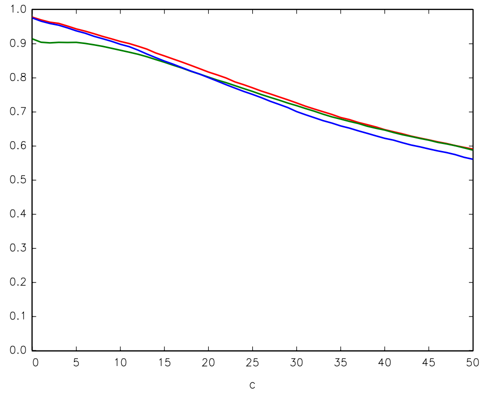


(p) $g = 20$, $\rho_{xy} = 0.3$

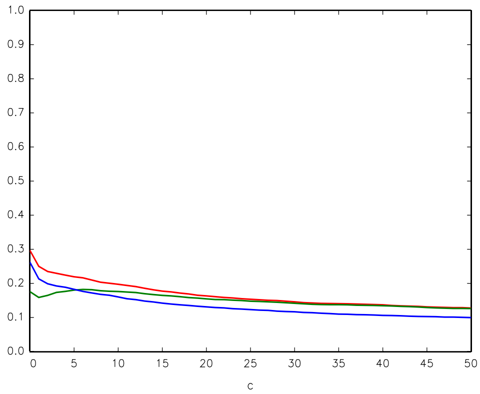
Figure S2 (continued). Asymptotic local power ($g = 5$ and $g = 20$) of nominal 0.05-level tests, $\mathcal{T}_{con}$: —— , $\mathcal{T}'_{con}$: —— , $\mathcal{T}_N$: ——



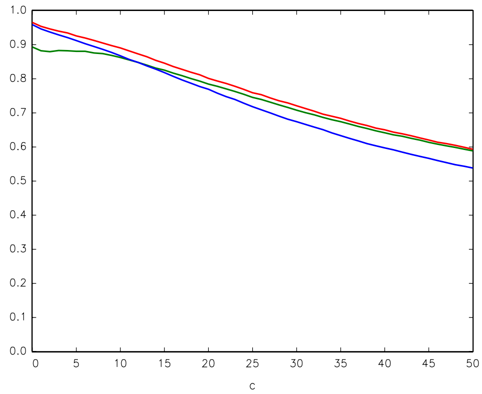
(q) $g = 5$, $\rho_{xy} = 0.5$



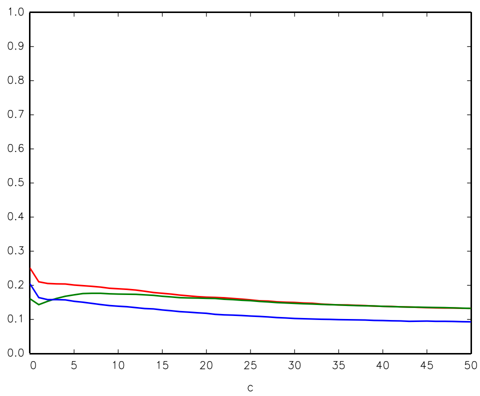
(r) $g = 20$, $\rho_{xy} = 0.5$



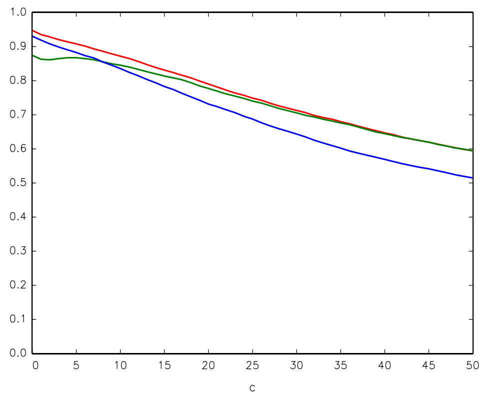
(s) $g = 5$, $\rho_{xy} = 0.7$



(t) $g = 20$, $\rho_{xy} = 0.7$

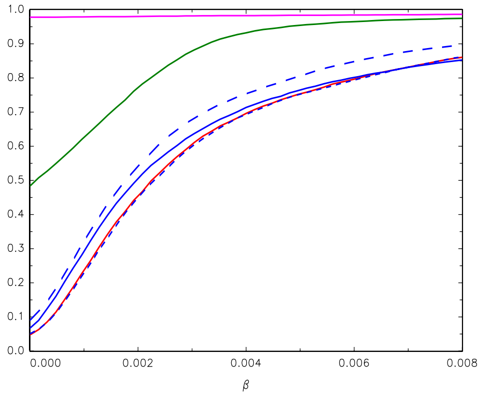


(u) $g = 5$, $\rho_{xy} = 0.9$

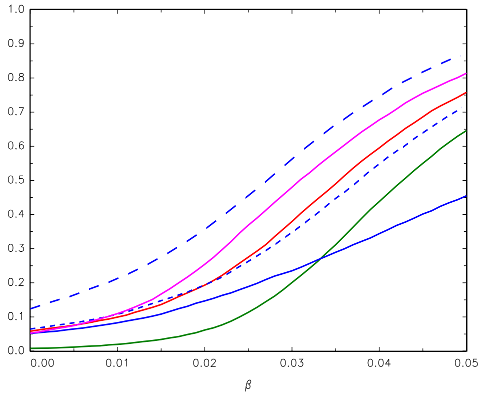


(v) $g = 20$, $\rho_{xy} = 0.9$

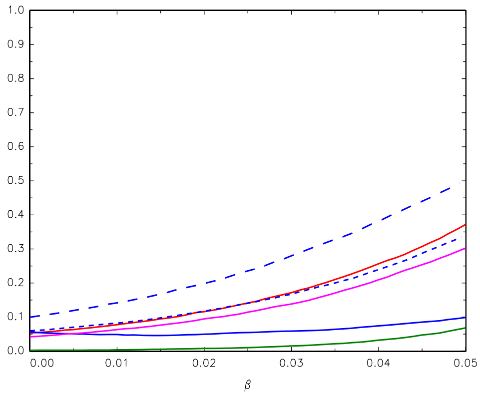
Figure S2 (continued). Asymptotic local power ($g = 5$ and $g = 20$) of nominal 0.05-level tests, $\mathcal{T}_{con}$: red line, $\mathcal{T}'_{con}$: green line, $\mathcal{T}_N$: blue line



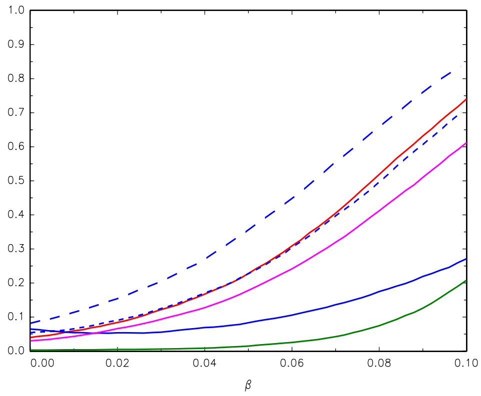
(a) $\phi = 1.025$



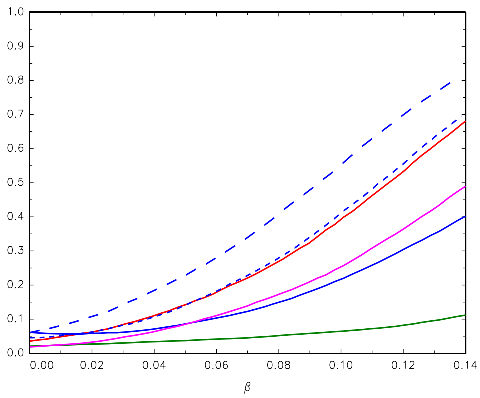
(b) $\phi = 1$



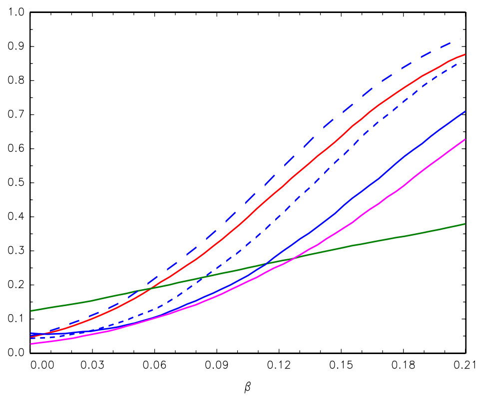
(c) $\phi = 0.975$



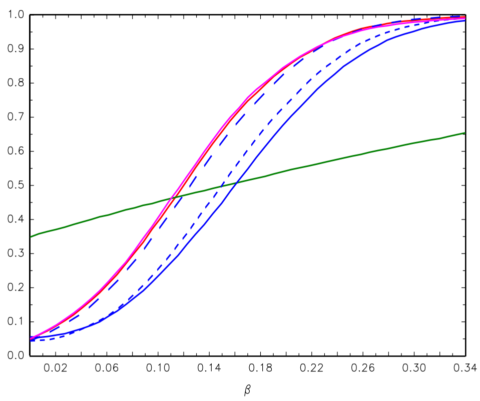
(d) $\phi = 0.95$



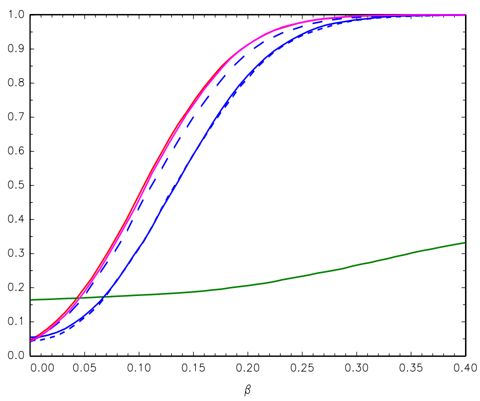
(e) $\phi = 0.875$



(f) $\phi = 0.75$

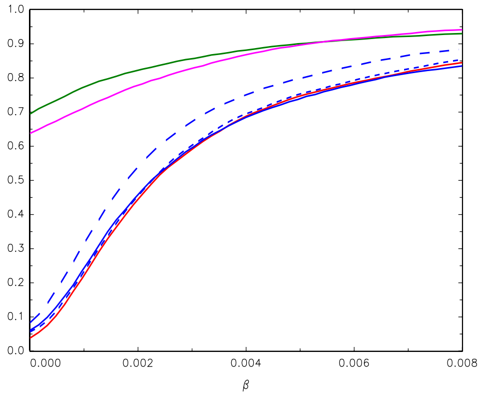


(g) $\phi = 0.5$

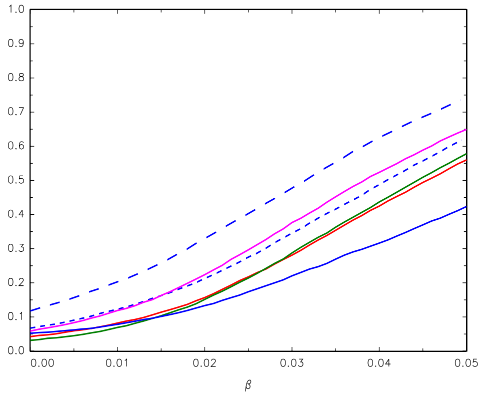


(h) $\phi = 0$

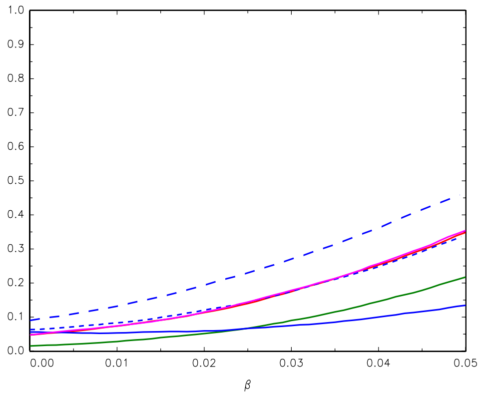
Figure S3. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.9$, $\theta = 0.5$; $\mathcal{T}_{hyb}$: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line, EMW: magenta solid line



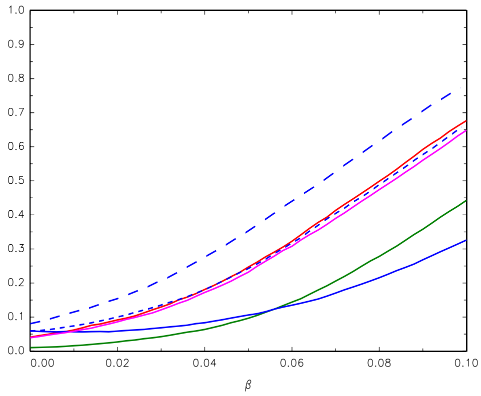
(a) $\phi = 1.025$



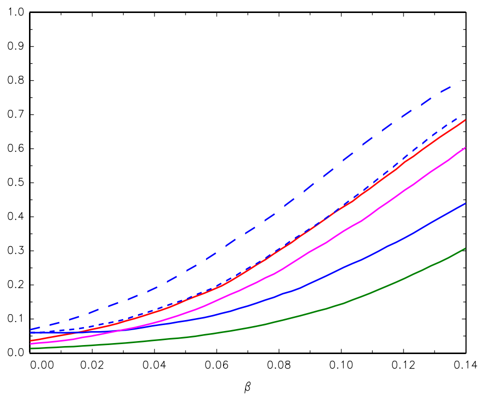
(b) $\phi = 1$



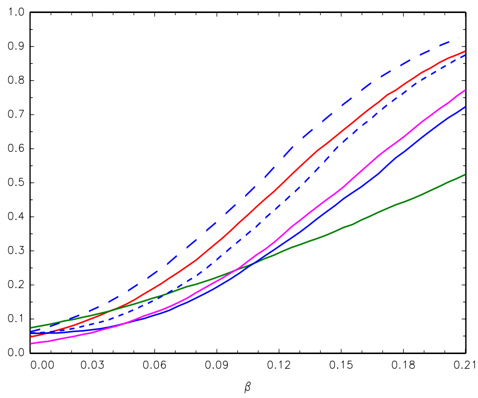
(c) $\phi = 0.975$



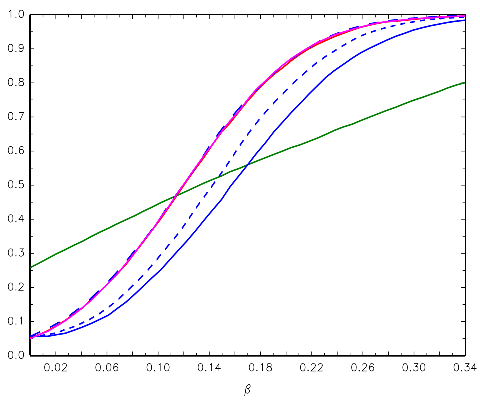
(d) $\phi = 0.95$



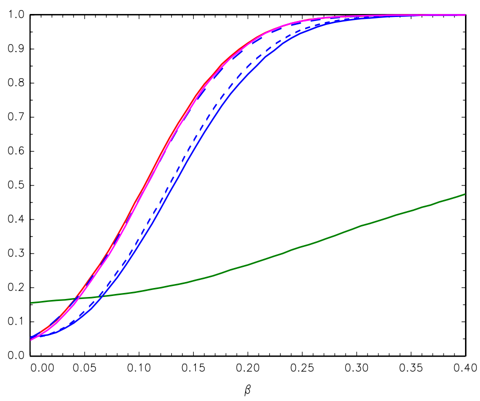
(e) $\phi = 0.875$



(f) $\phi = 0.75$

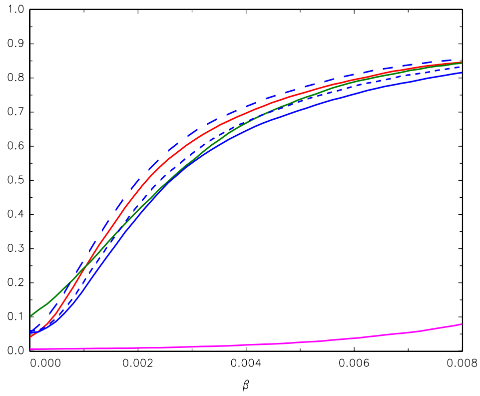


(g) $\phi = 0.5$

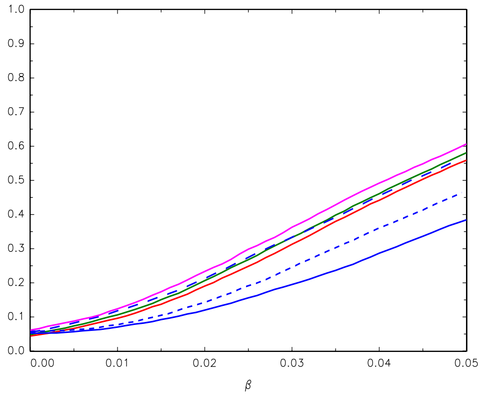


(h) $\phi = 0$

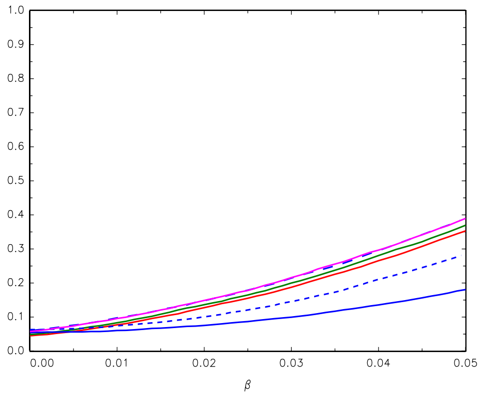
Figure S4. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.5$, $\theta = 0.5$; $\mathcal{T}_{hyb}$: , Q: , BD: , IVX_1: , IVX_2: , EMW:



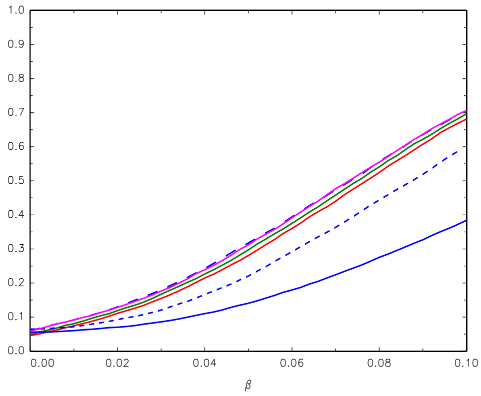
(a) $\phi = 1.025$



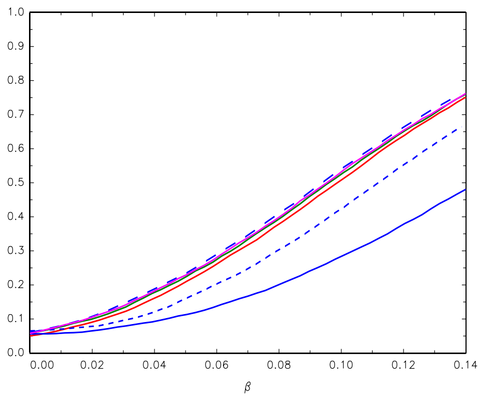
(b) $\phi = 1$



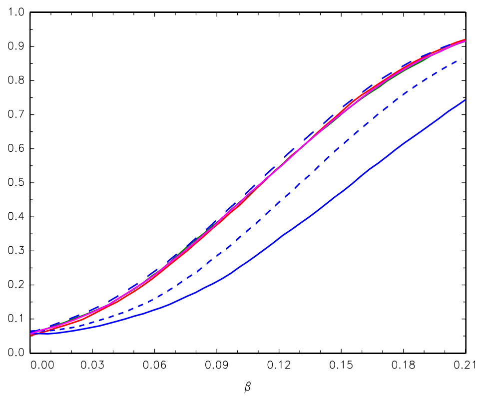
(c) $\phi = 0.975$



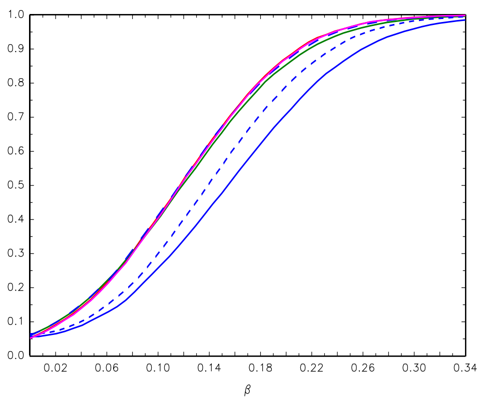
(d) $\phi = 0.95$



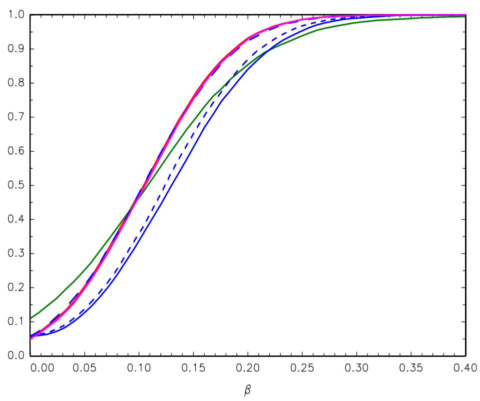
(e) $\phi = 0.875$



(f) $\phi = 0.75$

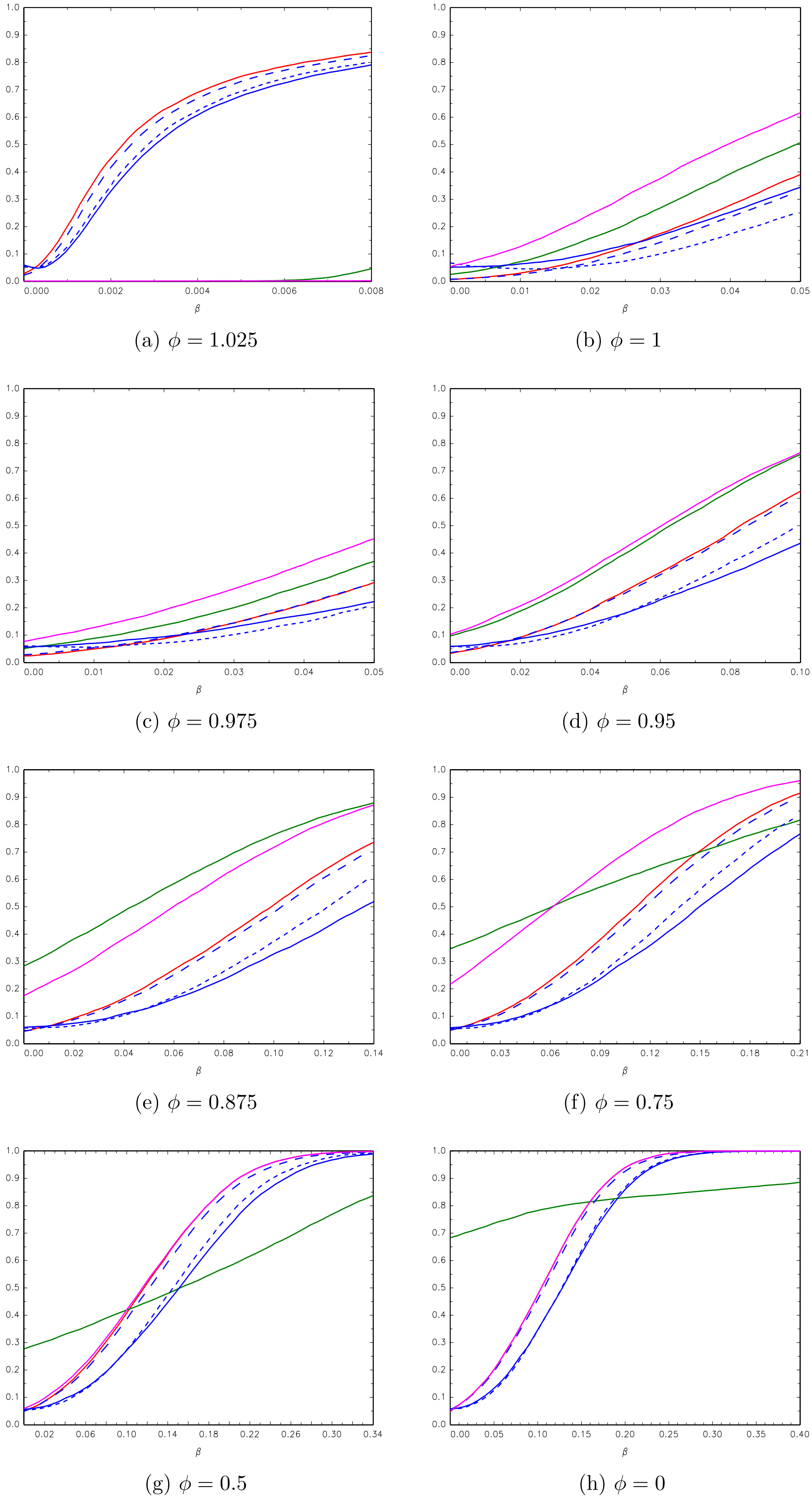


(g) $\phi = 0.5$



(h) $\phi = 0$

Figure S5. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0$, $\theta = 0.5$; $\mathcal{T}_{hyb}$: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line, EMW: magenta solid line



(a) $\phi = 1.025$

(b) $\phi = 1$

(c) $\phi = 0.975$

(d) $\phi = 0.95$

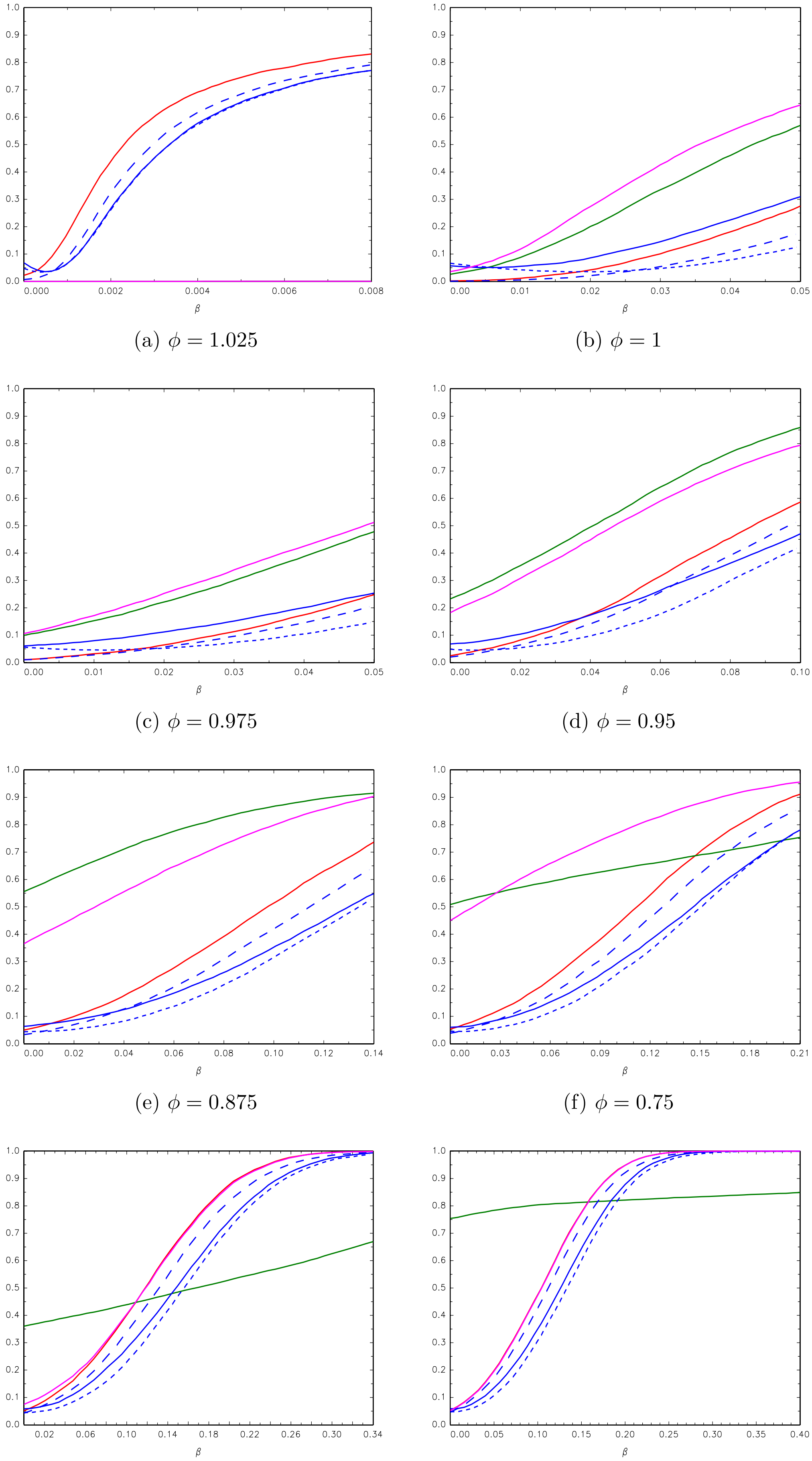
(e) $\phi = 0.875$

(f) $\phi = 0.75$

(g) $\phi = 0.5$

(h) $\phi = 0$

Figure S6. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.5$, $\theta = 0.5$; $\mathcal{T}_{hyb}$: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line, EMW: magenta solid line



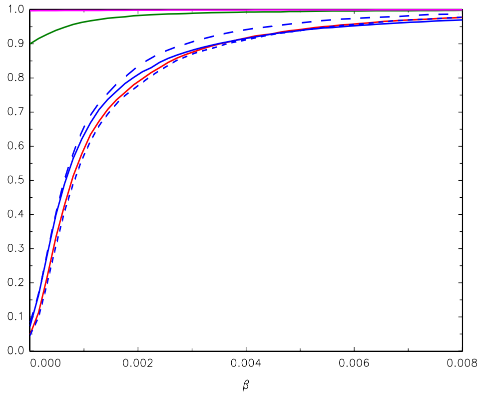
(a) $\phi = 1.025$ (b) $\phi = 1$

(c) $\phi = 0.975$ (d) $\phi = 0.95$

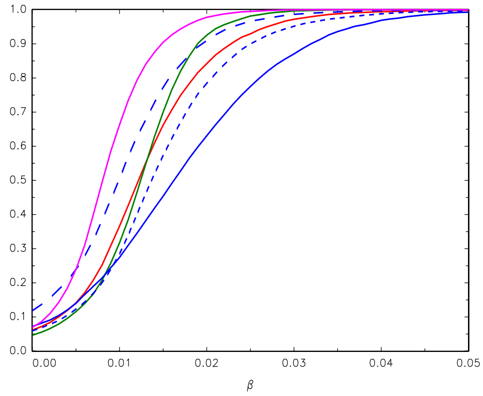
(e) $\phi = 0.875$ (f) $\phi = 0.75$

(g) $\phi = 0.5$ (h) $\phi = 0$

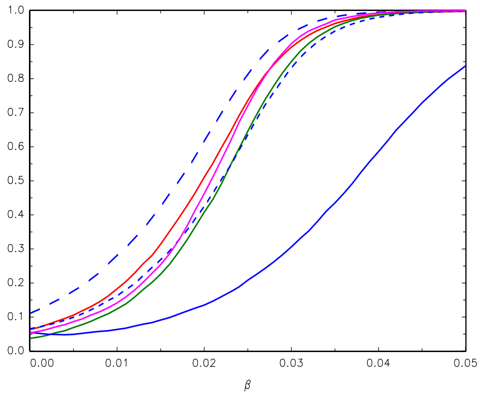
Figure S7. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.9$, $\theta = 0.5$; $\mathcal{T}_{hyb}$: ——, Q: ——, BD: ——, IVX_1: — —, IVX_2: - - - -, EMW: ——



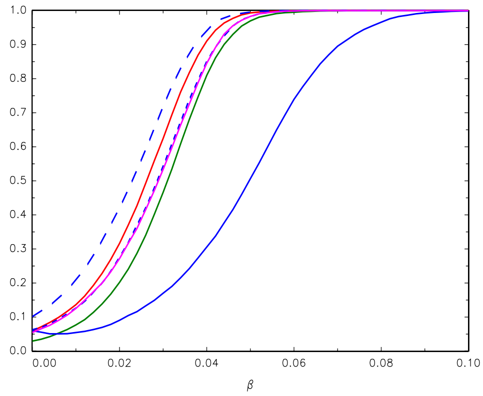
(a) $\phi = 1.025$



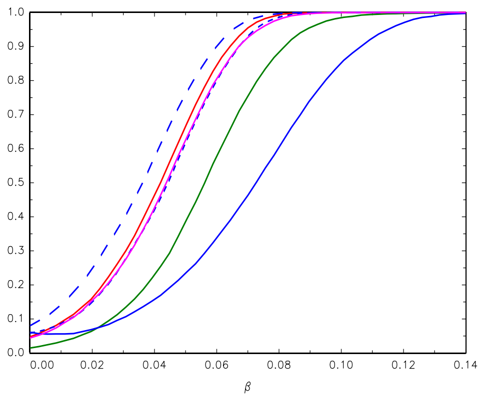
(b) $\phi = 1$



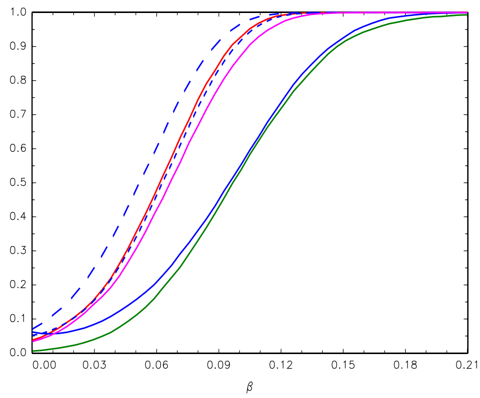
(c) $\phi = 0.975$



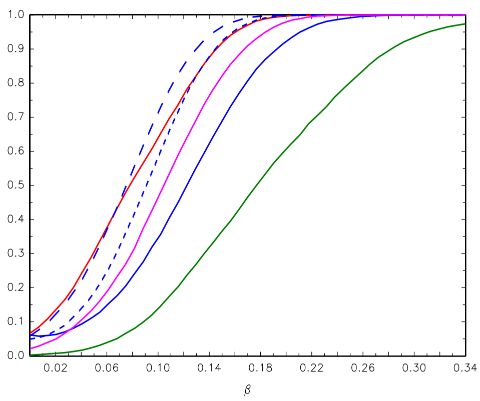
(d) $\phi = 0.95$



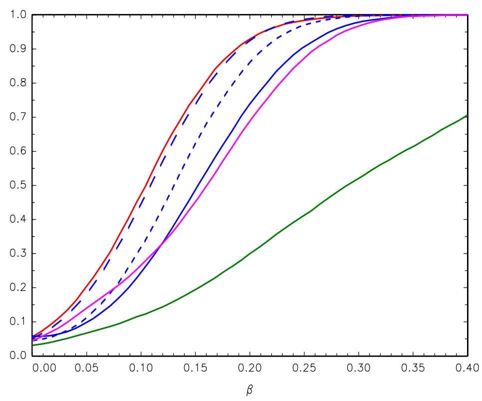
(e) $\phi = 0.875$



(f) $\phi = 0.75$

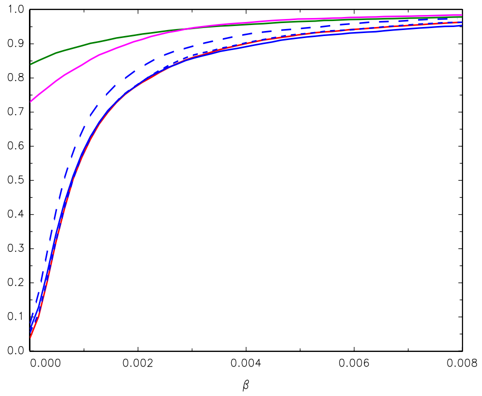


(g) $\phi = 0.5$

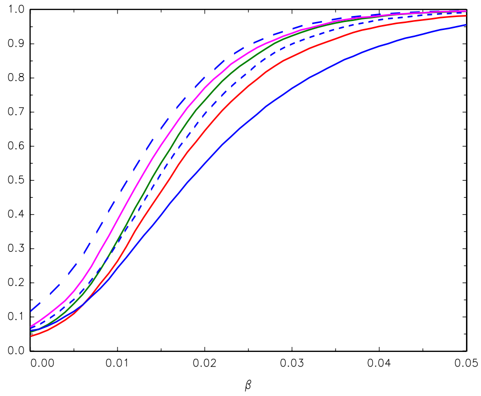


(h) $\phi = 0$

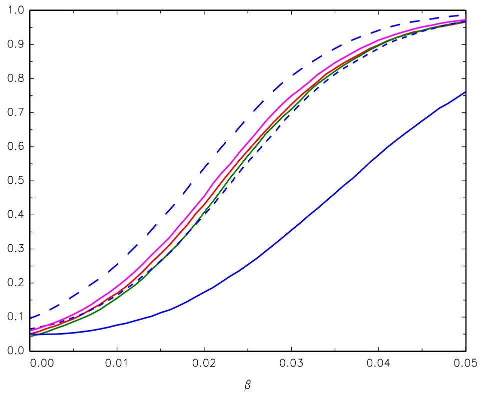
Figure S8. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.9$, $\theta = -0.5$; $\mathcal{T}_{hyb}$: —— , Q: —— , BD: —— , IVX_1: – – – , IVX_2: - - - - , EMW: ——



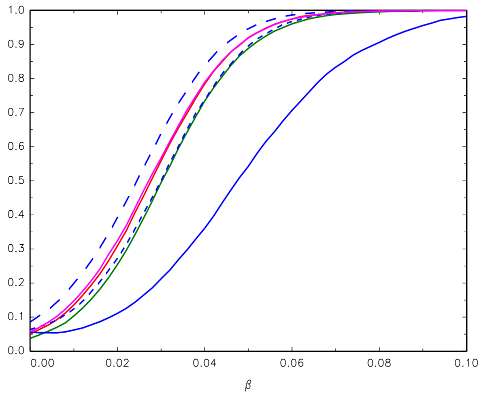
(a) $\phi = 1.025$



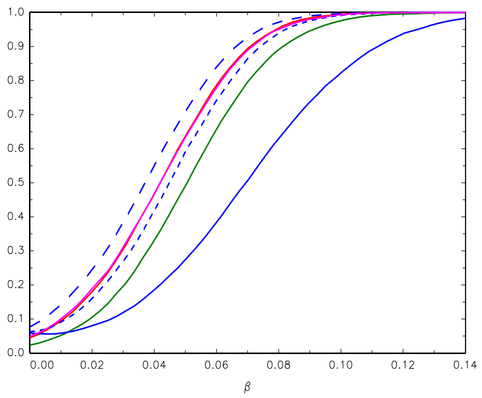
(b) $\phi = 1$



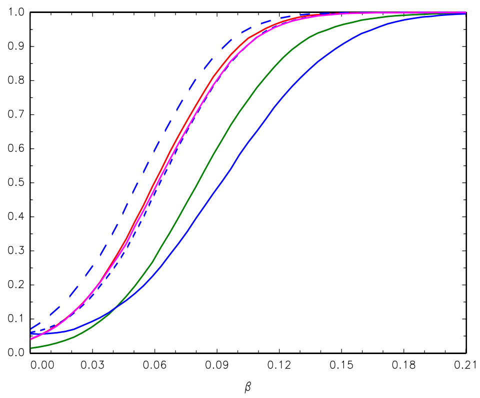
(c) $\phi = 0.975$



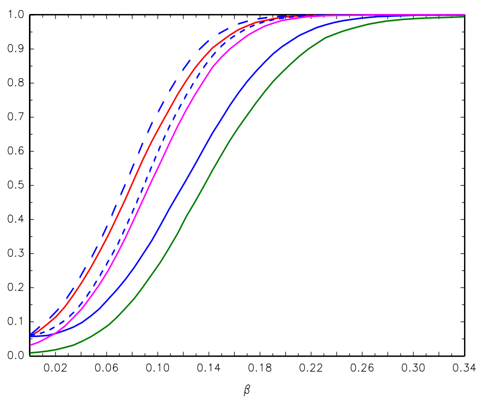
(d) $\phi = 0.95$



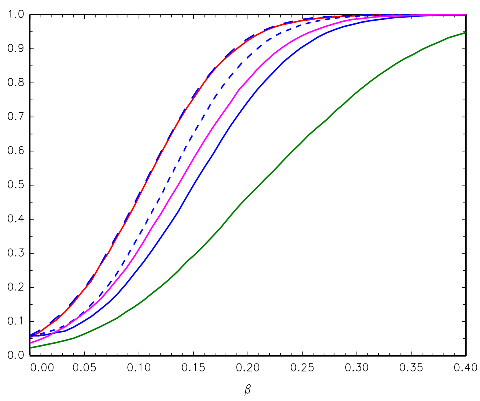
(e) $\phi = 0.875$



(f) $\phi = 0.75$

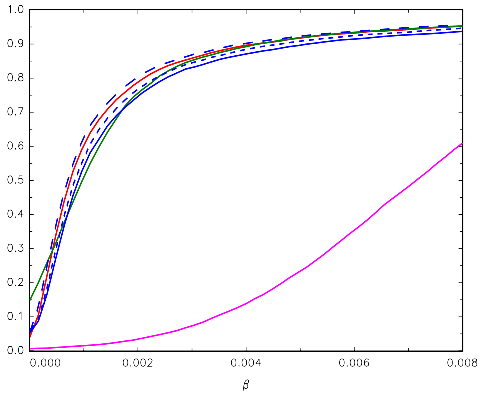


(g) $\phi = 0.5$

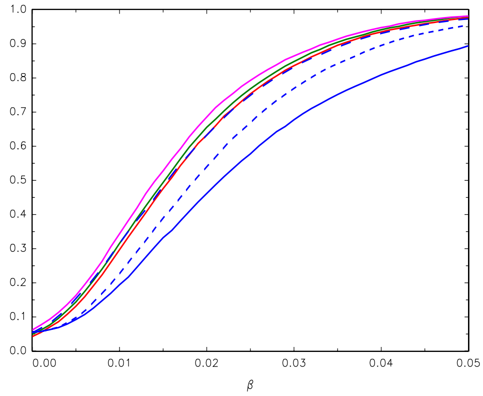


(h) $\phi = 0$

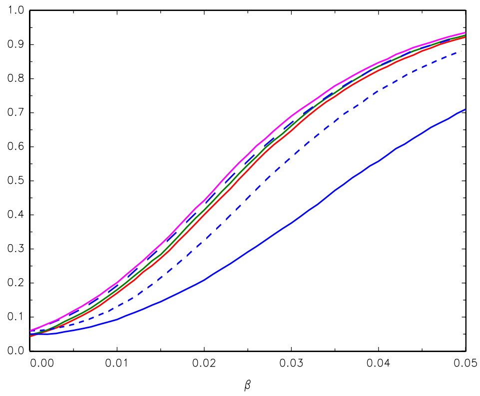
Figure S9. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.5$, $\theta = -0.5$; $\mathcal{T}_{hyb}$: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line, EMW: magenta solid line



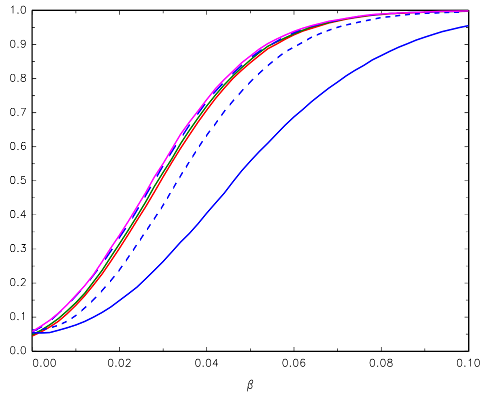
(a) $\phi = 1.025$



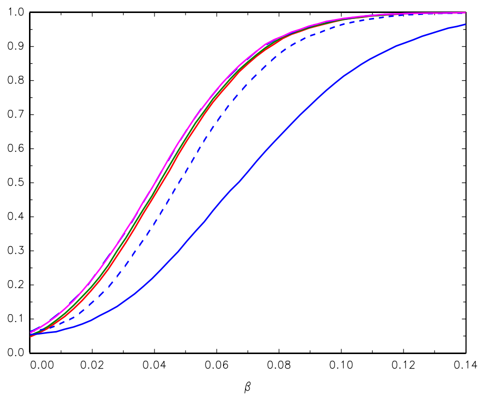
(b) $\phi = 1$



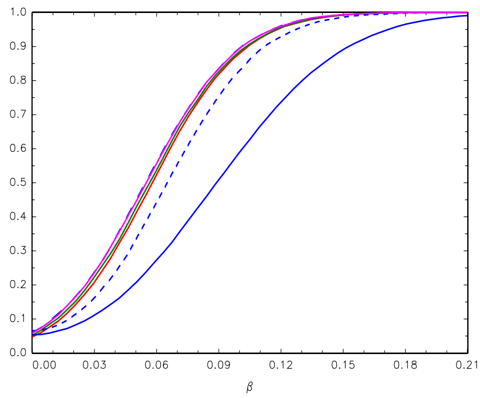
(c) $\phi = 0.975$



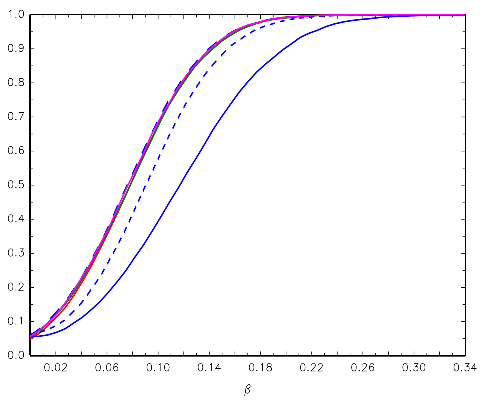
(d) $\phi = 0.95$



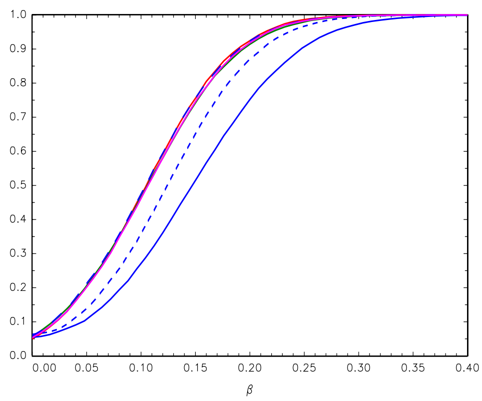
(e) $\phi = 0.875$



(f) $\phi = 0.75$

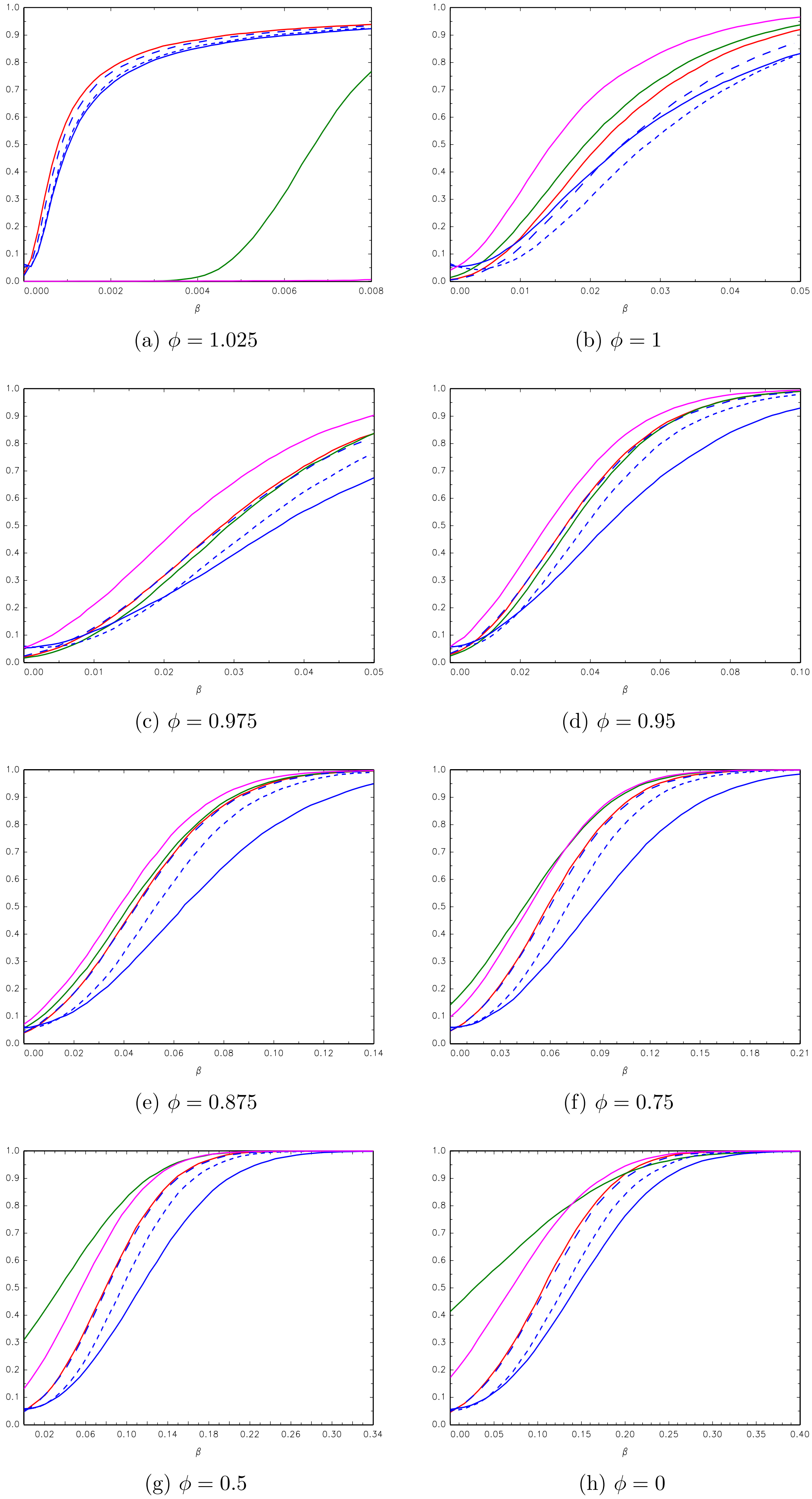


(g) $\phi = 0.5$



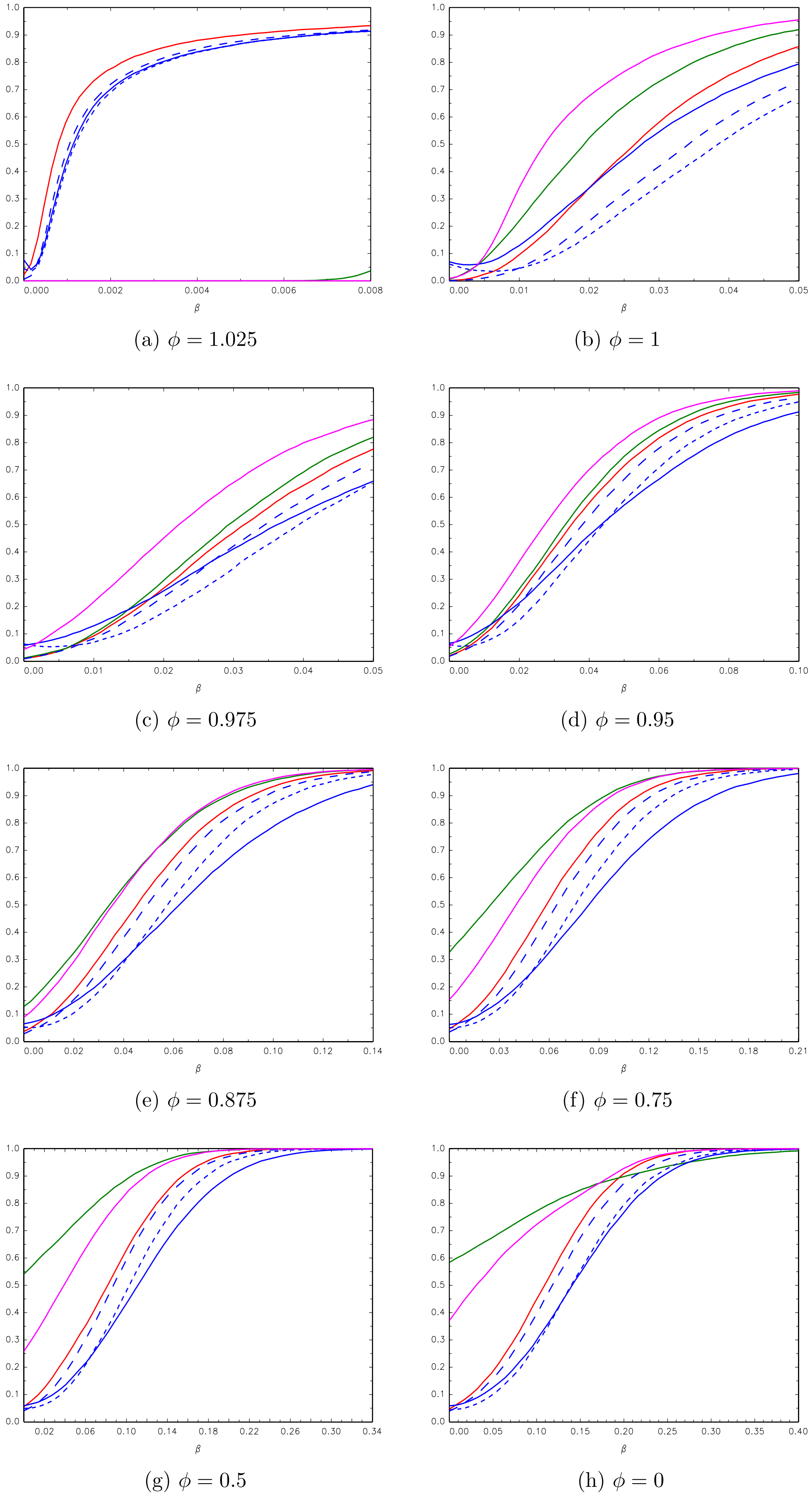
(h) $\phi = 0$

Figure S10. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0$, $\theta = -0.5$; $\mathcal{T}_{hyb}$: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line, EMW: magenta solid line



(a) $\phi = 1.025$ (b) $\phi = 1$ (c) $\phi = 0.975$ (d) $\phi = 0.95$ (e) $\phi = 0.875$ (f) $\phi = 0.75$ (g) $\phi = 0.5$ (h) $\phi = 0$

Figure S11. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.5$, $\theta = -0.5$; $\mathcal{T}_{hyb}$: —— , Q: —— , BD: —— , IVX_1: – – , IVX_2: - - - , EMW: ——



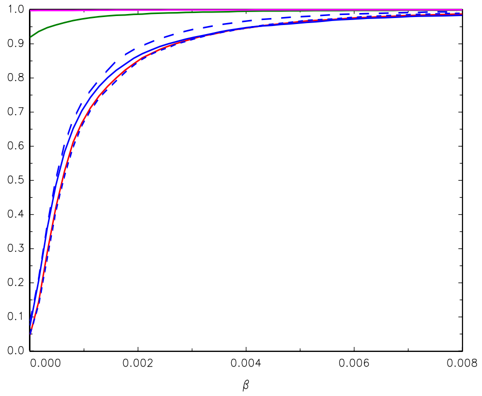
(a) $\phi = 1.025$ (b) $\phi = 1$

(c) $\phi = 0.975$ (d) $\phi = 0.95$

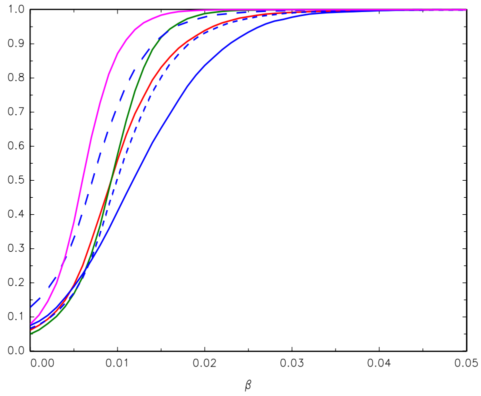
(e) $\phi = 0.875$ (f) $\phi = 0.75$

(g) $\phi = 0.5$ (h) $\phi = 0$

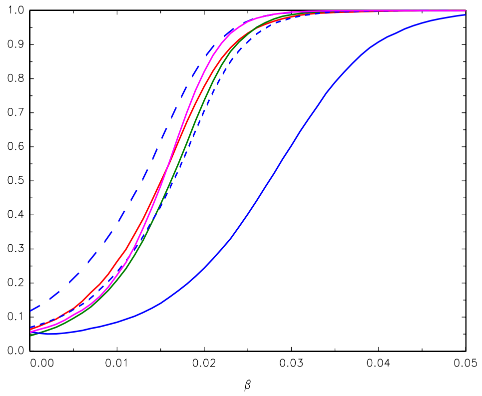
Figure S12. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.9$, $\theta = -0.5$; $\mathcal{T}_{hyb}$: ——, Q: ——, BD: ——, IVX_1: — —, IVX_2: - - -, EMW: ——



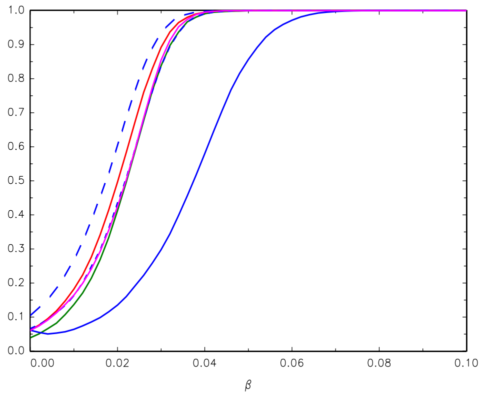
(a) $\phi = 1.025$



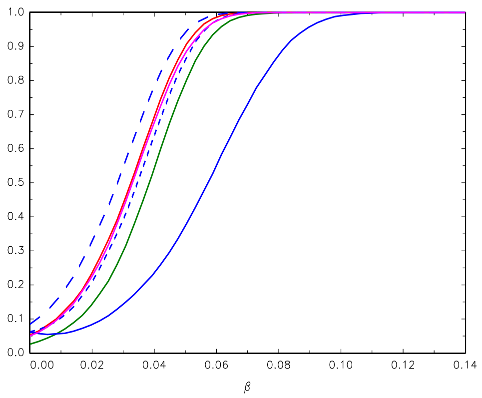
(b) $\phi = 1$



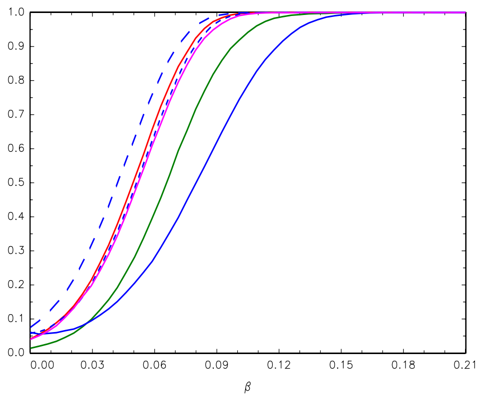
(c) $\phi = 0.975$



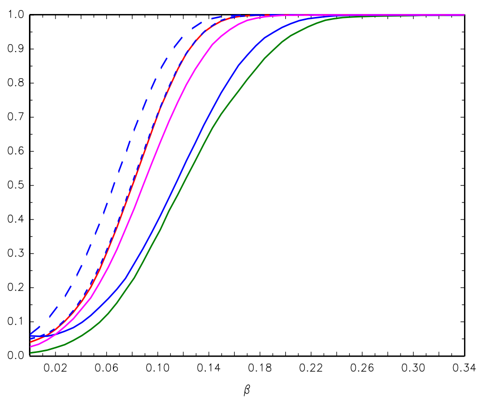
(d) $\phi = 0.95$



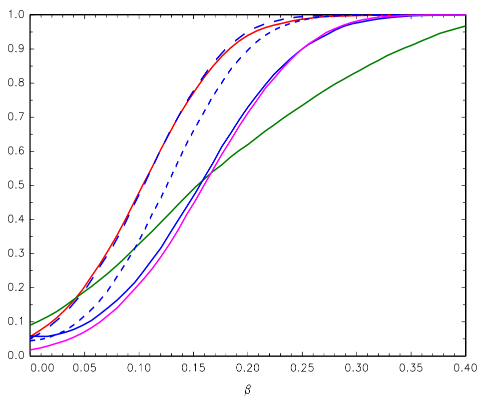
(e) $\phi = 0.875$



(f) $\phi = 0.75$

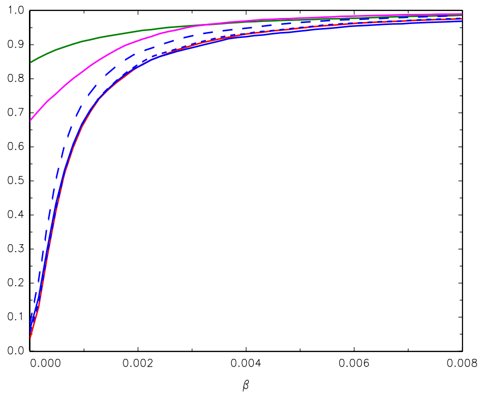


(g) $\phi = 0.5$

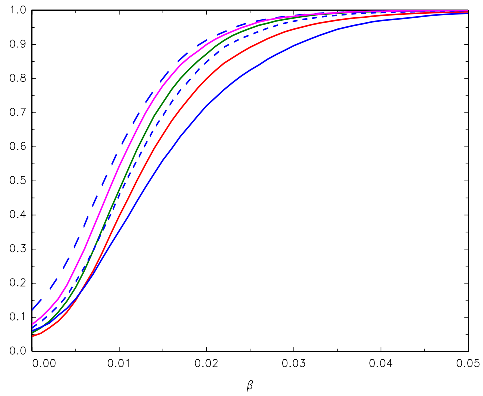


(h) $\phi = 0$

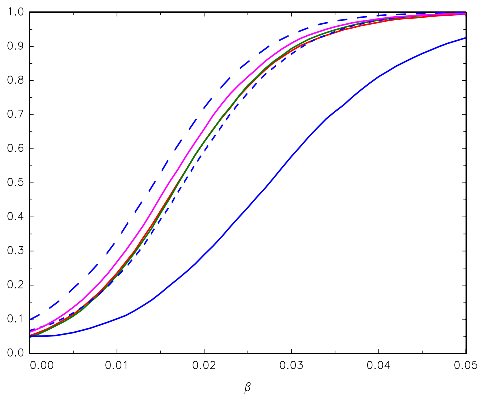
Figure S13. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.9$, $\delta = 0.5$; $\mathcal{T}_{hyb}$: ——, Q: ——, BD: ——, IVX_1: – – –, IVX_2: - - - -, EMW: ——



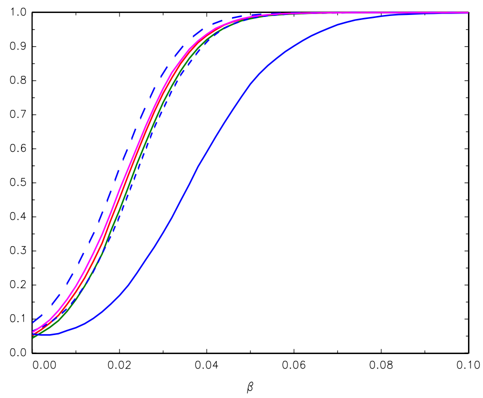
(a) $\phi = 1.025$



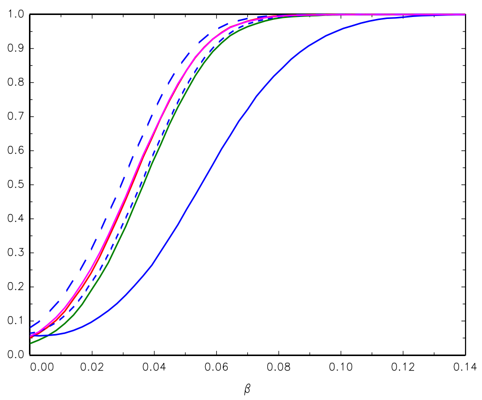
(b) $\phi = 1$



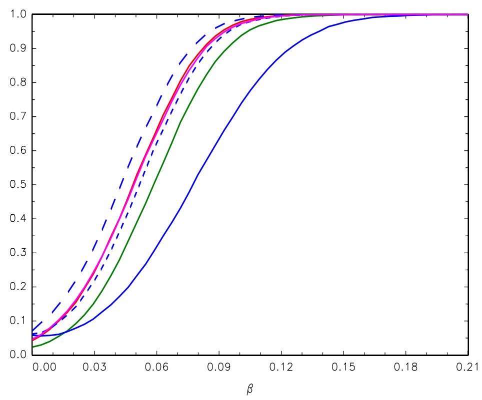
(c) $\phi = 0.975$



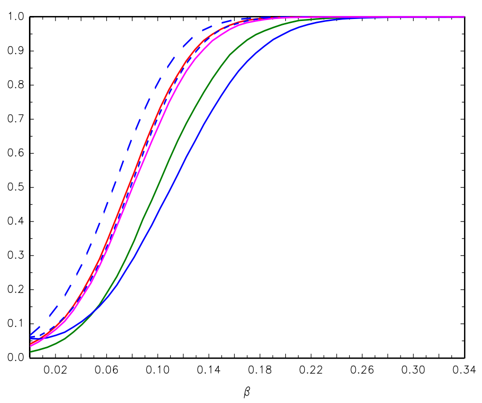
(d) $\phi = 0.95$



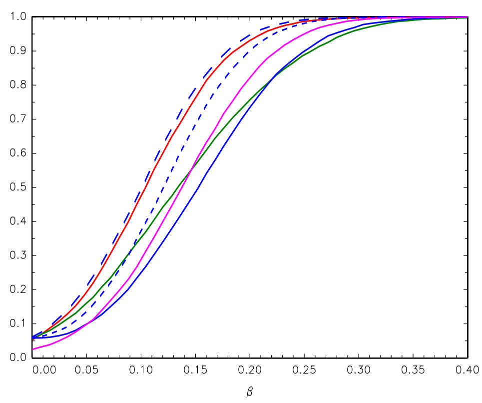
(e) $\phi = 0.875$



(f) $\phi = 0.75$

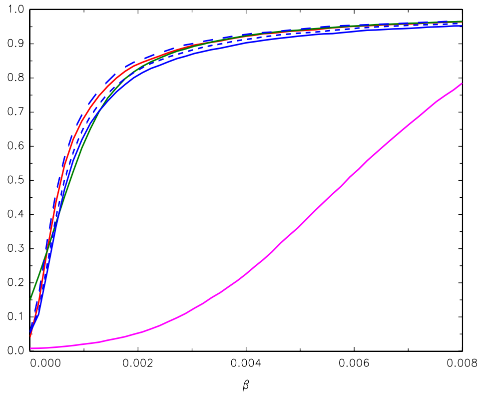


(g) $\phi = 0.5$

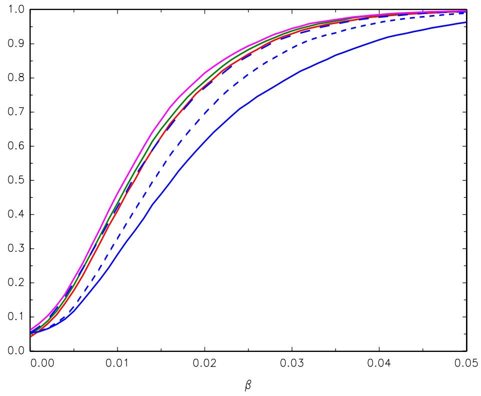


(h) $\phi = 0$

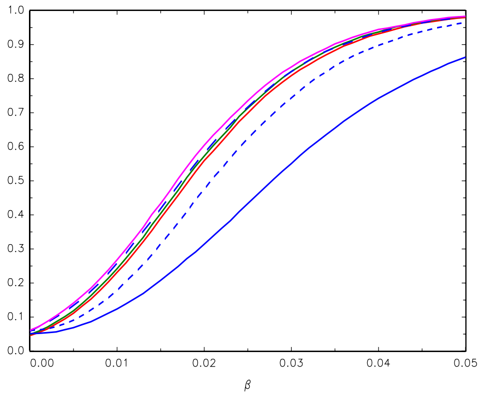
Figure S14. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.5$, $\delta = 0.5$; $\mathcal{T}_{hyb}$: —— , Q: —— , BD: —— , IVX_1: – – , IVX_2: - - - , EMW: ——



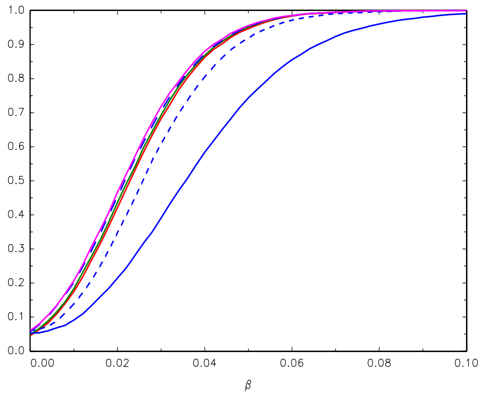
(a) $\phi = 1.025$



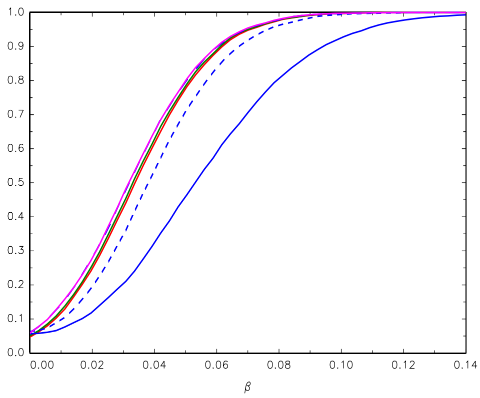
(b) $\phi = 1$



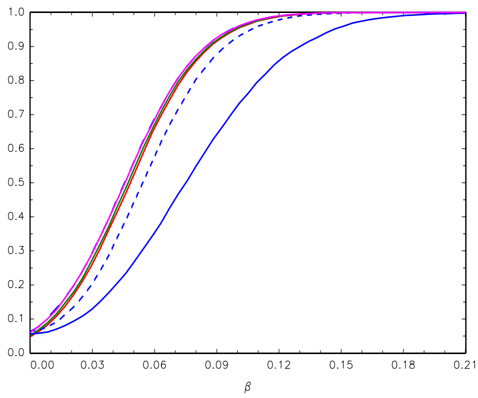
(c) $\phi = 0.975$



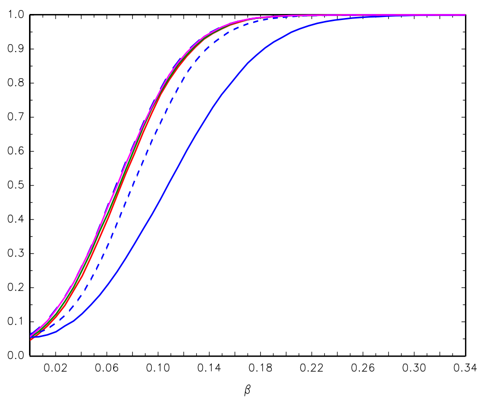
(d) $\phi = 0.95$



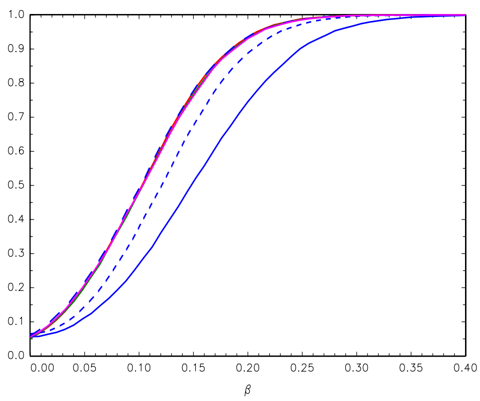
(e) $\phi = 0.875$



(f) $\phi = 0.75$

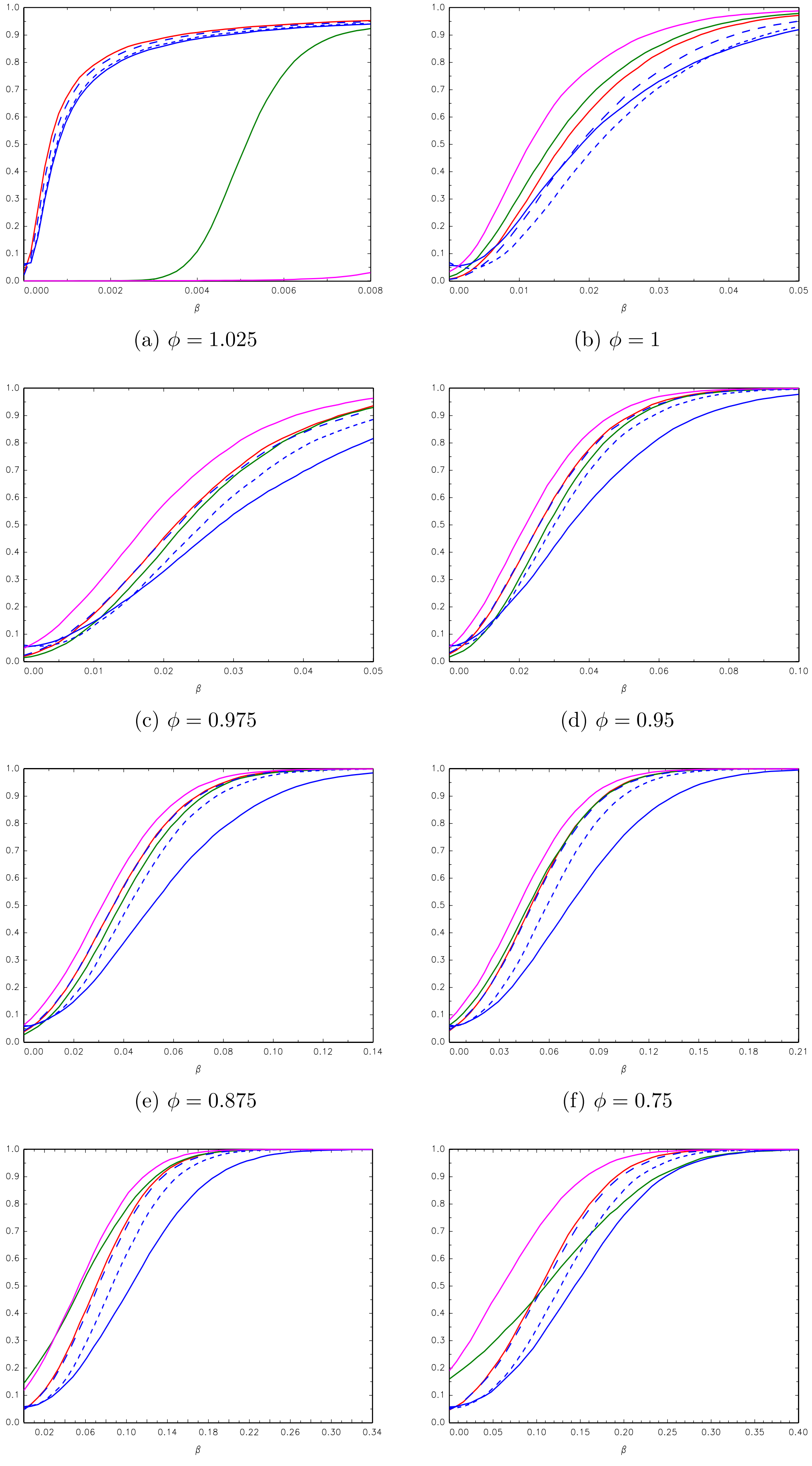


(g) $\phi = 0.5$



(h) $\phi = 0$

Figure S15. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0$, $\delta = 0.5$; $\mathcal{T}_{hyb}$: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line, EMW: magenta solid line



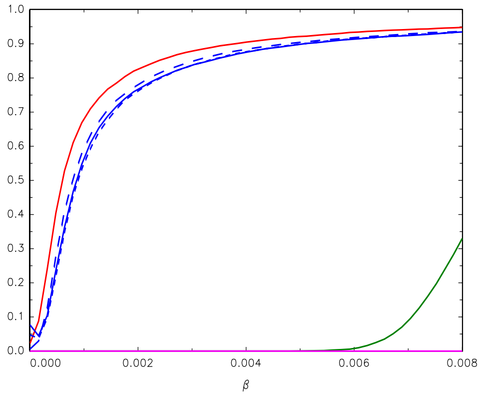
(a) $\phi = 1.025$ (b) $\phi = 1$

(c) $\phi = 0.975$ (d) $\phi = 0.95$

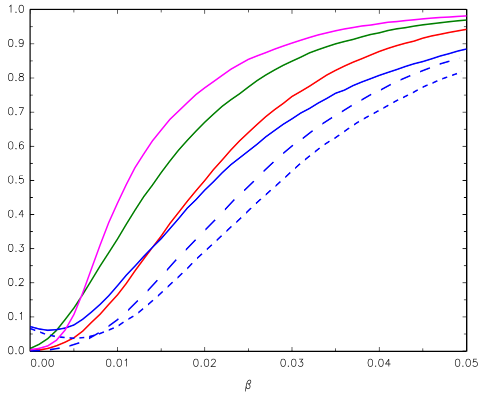
(e) $\phi = 0.875$ (f) $\phi = 0.75$

(g) $\phi = 0.5$ (h) $\phi = 0$

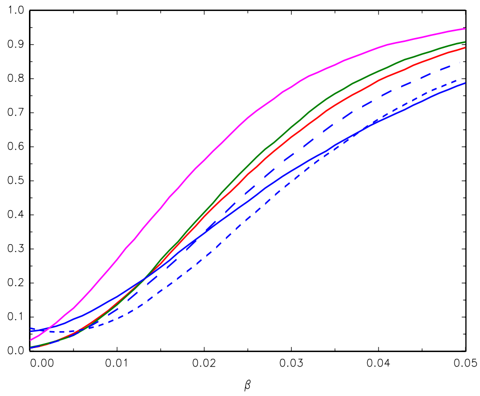
Figure S16. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.5$, $\delta = 0.5$; $\mathcal{T}_{hyb}$: red solid, Q: green solid, BD: blue solid, IVX_1: blue long-dashed, IVX_2: blue short-dashed, EMW: magenta solid



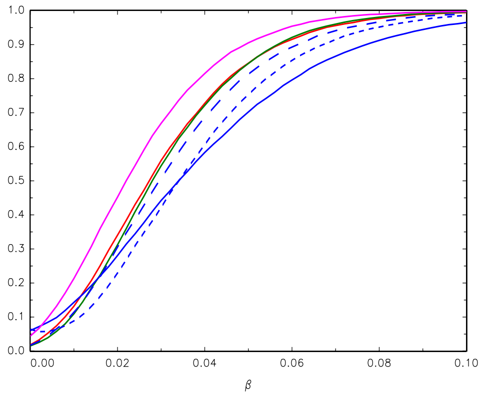
(a) $\phi = 1.025$



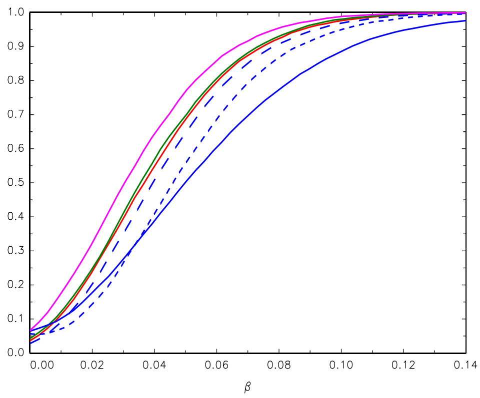
(b) $\phi = 1$



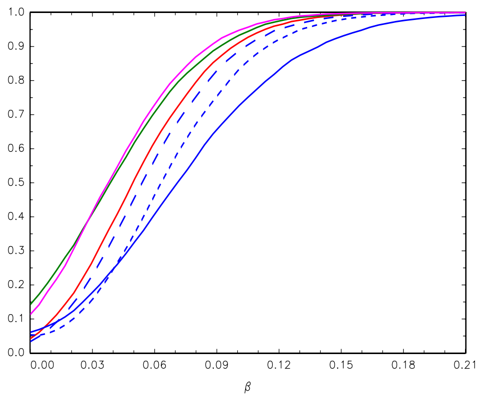
(c) $\phi = 0.975$



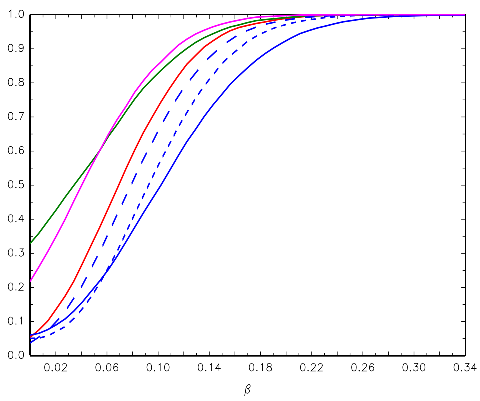
(d) $\phi = 0.95$



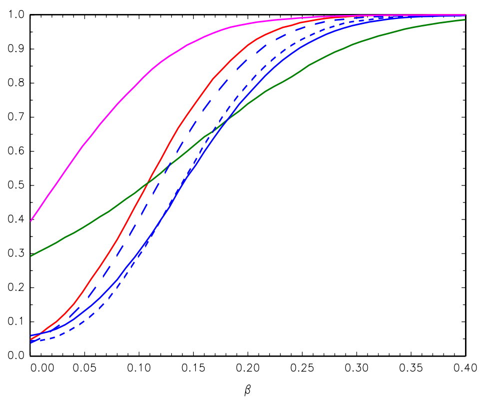
(e) $\phi = 0.875$



(f) $\phi = 0.75$

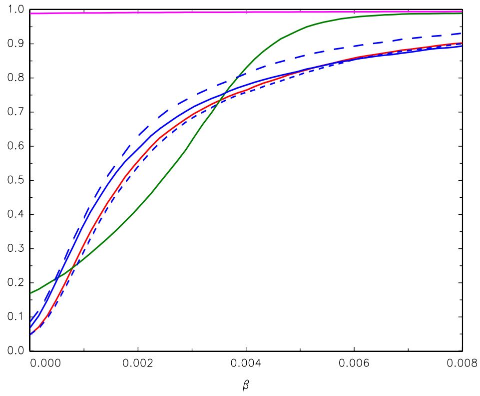


(g) $\phi = 0.5$

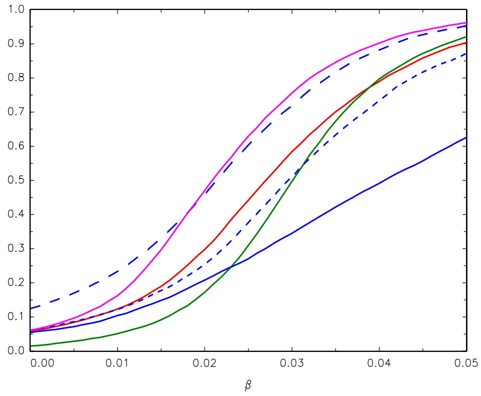


(h) $\phi = 0$

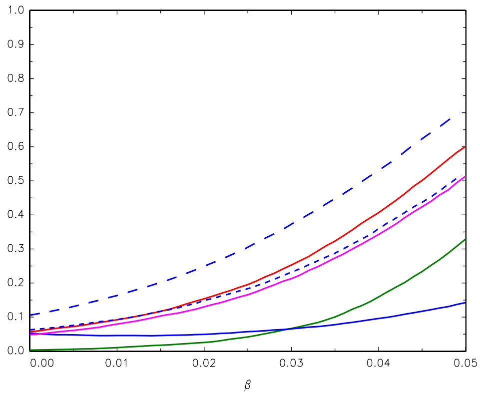
Figure S17. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.9$, $\delta = 0.5$; $\mathcal{T}_{hyb}$: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line, EMW: magenta solid line



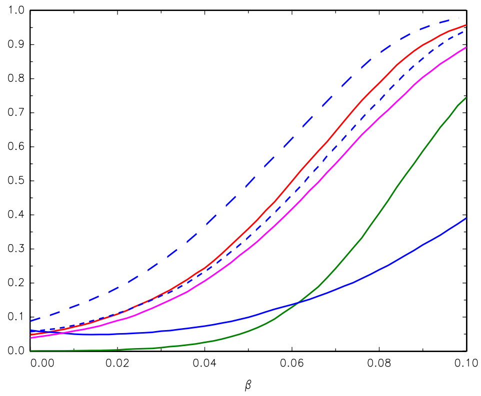
(a) $\phi = 1.025$



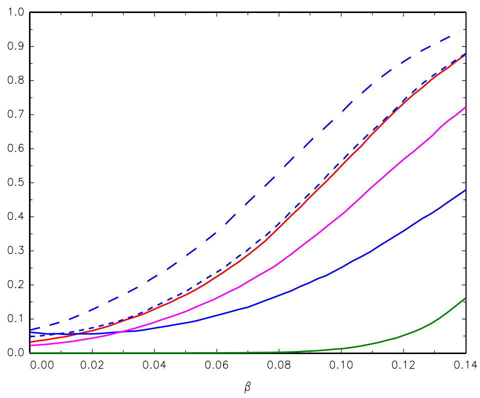
(b) $\phi = 1$



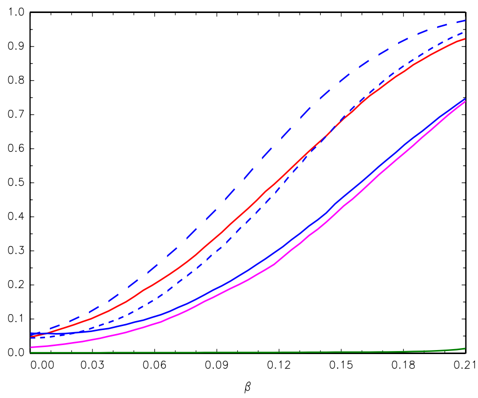
(c) $\phi = 0.975$



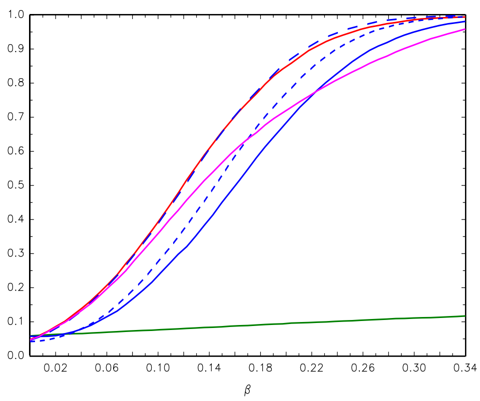
(d) $\phi = 0.95$



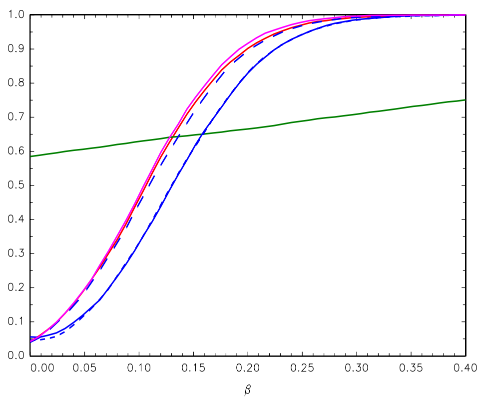
(e) $\phi = 0.875$



(f) $\phi = 0.75$

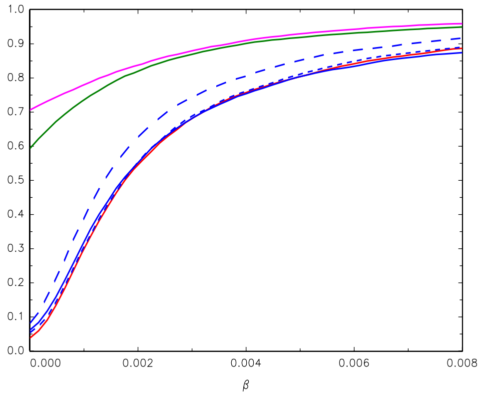


(g) $\phi = 0.5$

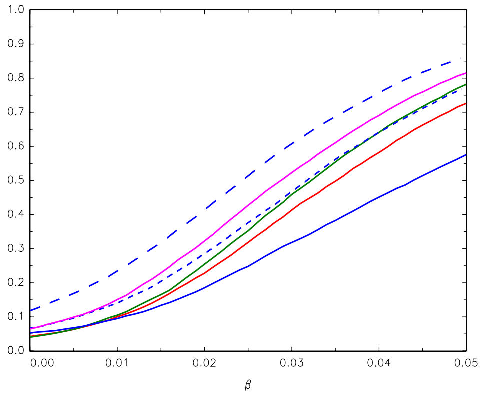


(h) $\phi = 0$

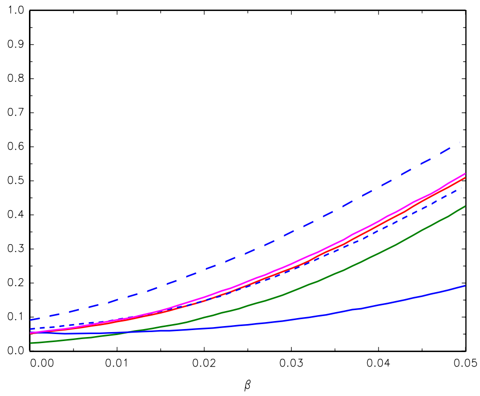
Figure S18. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.9$, $\delta = -0.5$; $\mathcal{T}_{hyb}$: ——, Q: ——, BD: ——, IVX_1: — —, IVX_2: - - -, EMW: ——



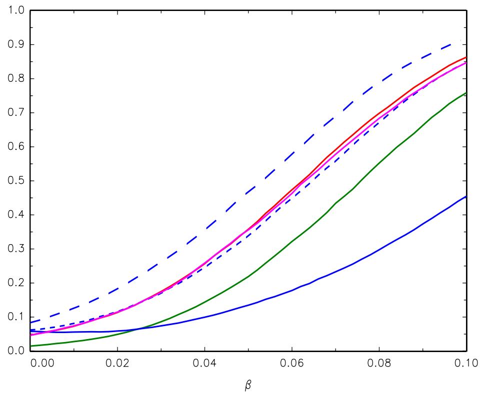
(a) $\phi = 1.025$



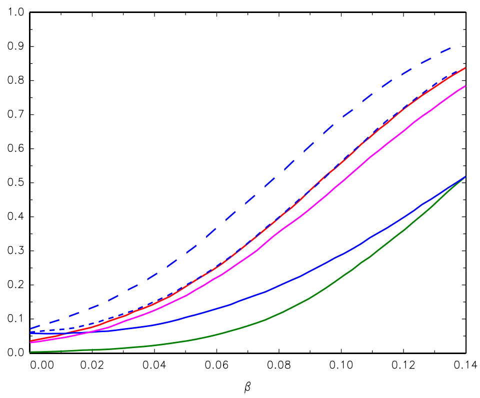
(b) $\phi = 1$



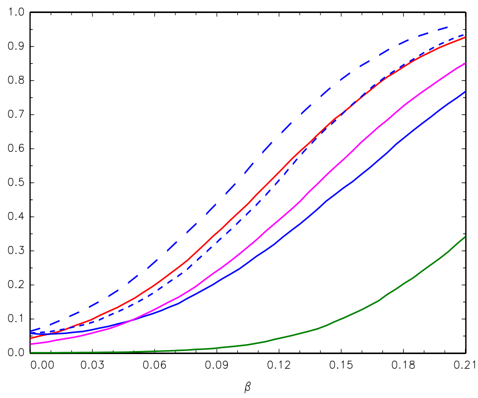
(c) $\phi = 0.975$



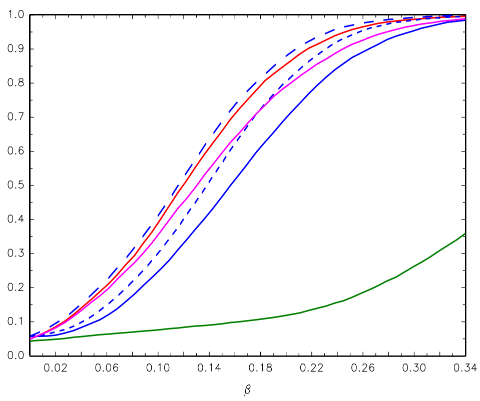
(d) $\phi = 0.95$



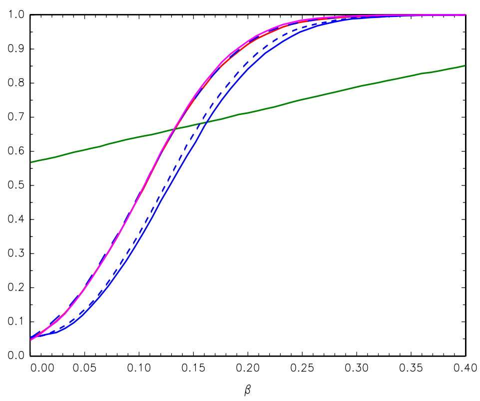
(e) $\phi = 0.875$



(f) $\phi = 0.75$

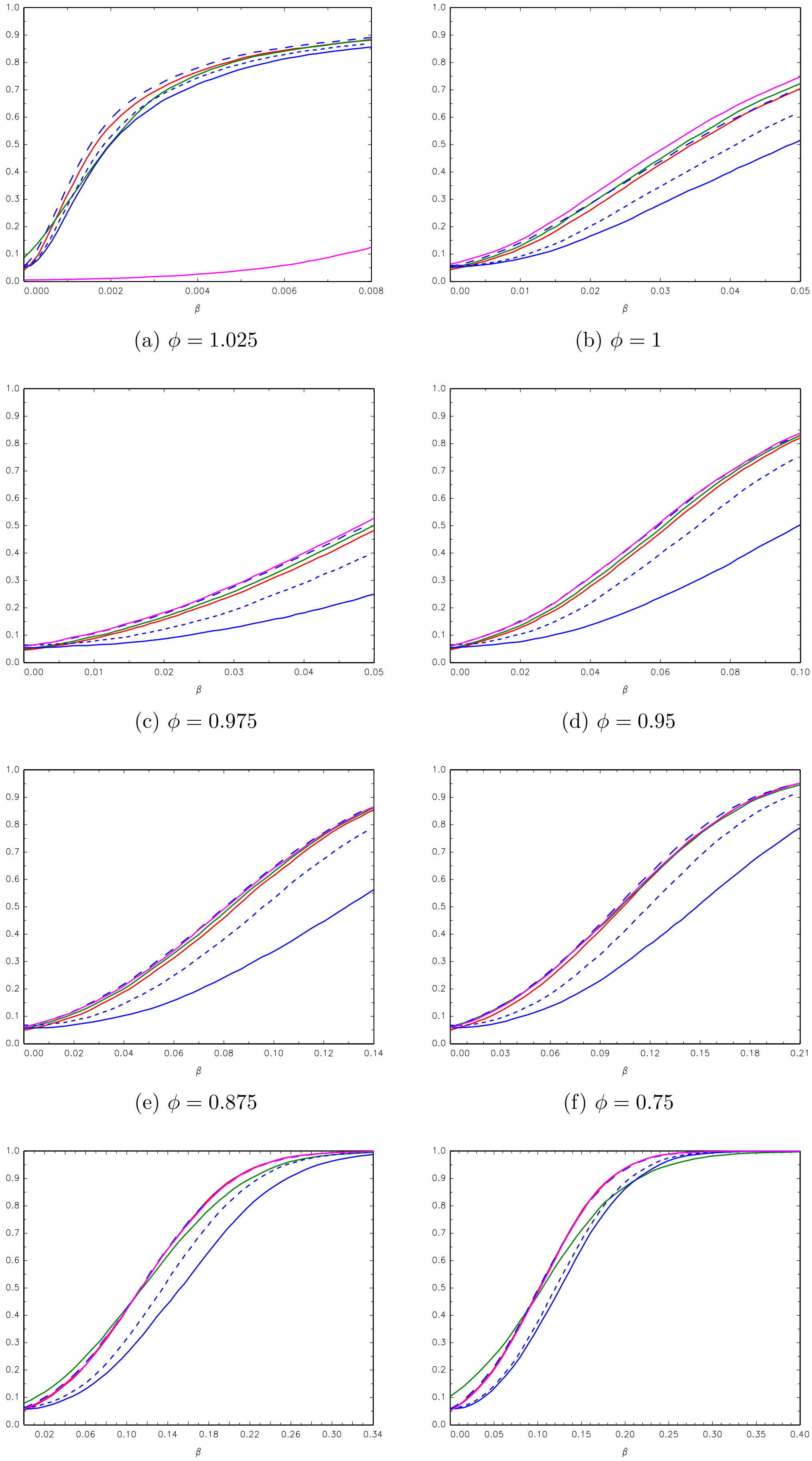


(g) $\phi = 0.5$



(h) $\phi = 0$

Figure S19. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.5$, $\delta = -0.5$; $\mathcal{T}_{hyb}$: —— , Q: —— , BD: —— , IVX_1: — — , IVX_2: - - - , EMW: ——



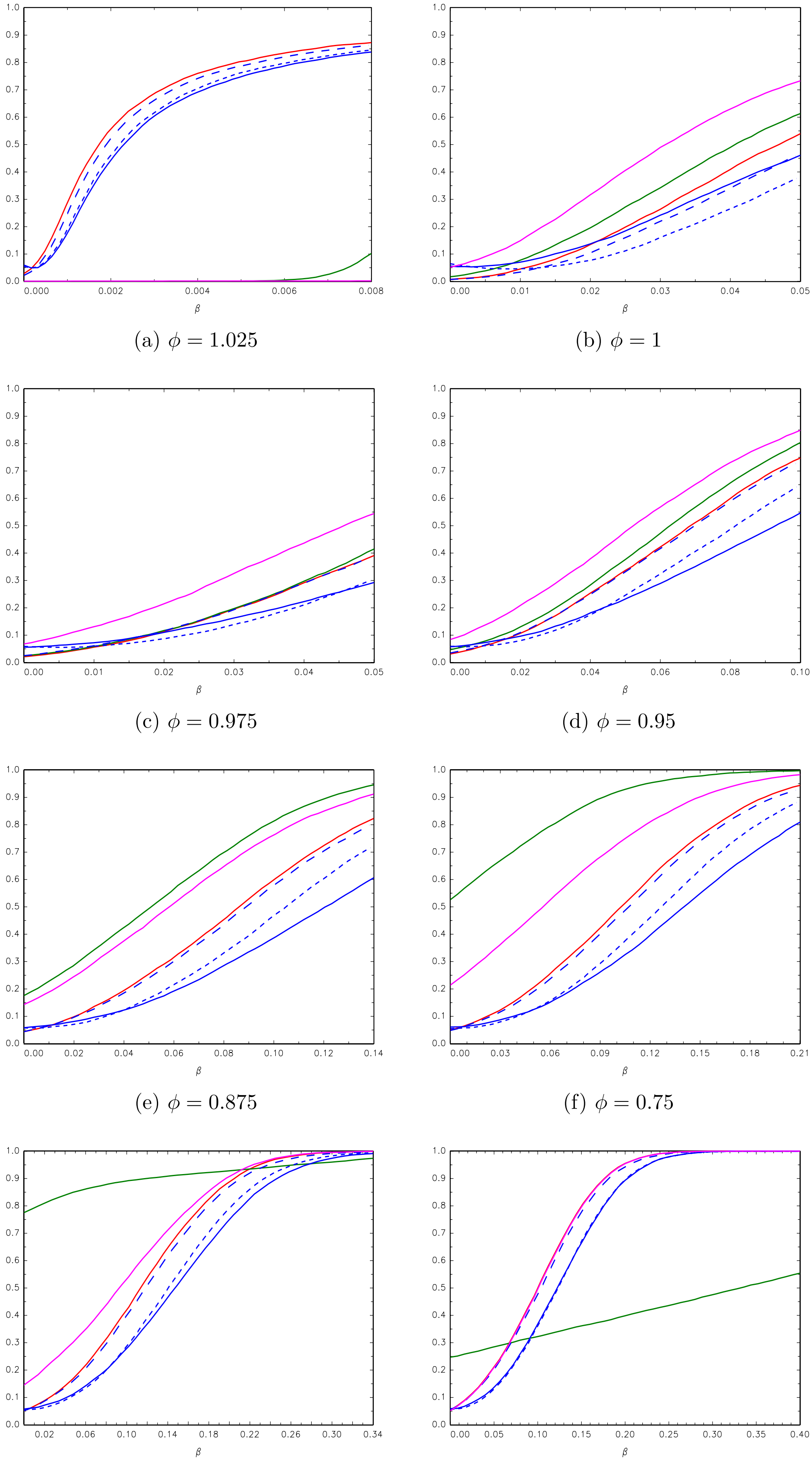
(a) $\phi = 1.025$ (b) $\phi = 1$

(c) $\phi = 0.975$ (d) $\phi = 0.95$

(e) $\phi = 0.875$ (f) $\phi = 0.75$

(g) $\phi = 0.5$ (h) $\phi = 0$

Figure S20. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0$, $\delta = -0.5$; $\mathcal{T}_{hyb}$: —— , Q: —— , BD: —— , IVX_1: – – – , IVX_2: - - - - , EMW: ——



(a) $\phi = 1.025$

(b) $\phi = 1$

(c) $\phi = 0.975$

(d) $\phi = 0.95$

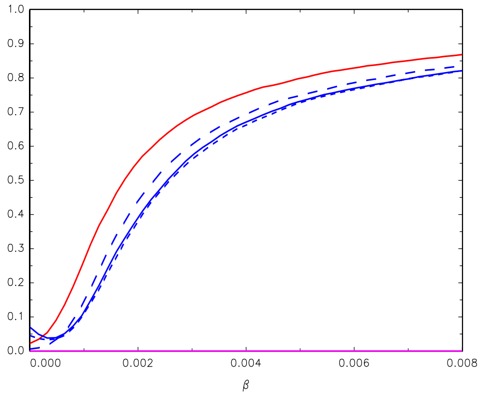
(e) $\phi = 0.875$

(f) $\phi = 0.75$

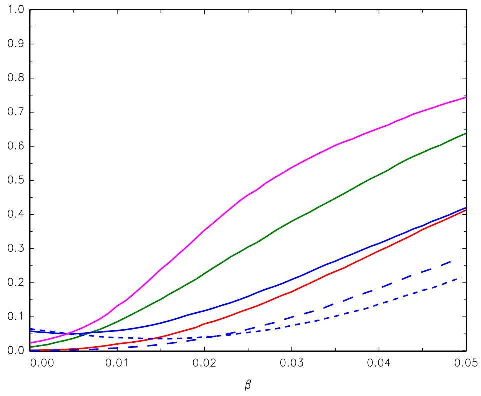
(g) $\phi = 0.5$

(h) $\phi = 0$

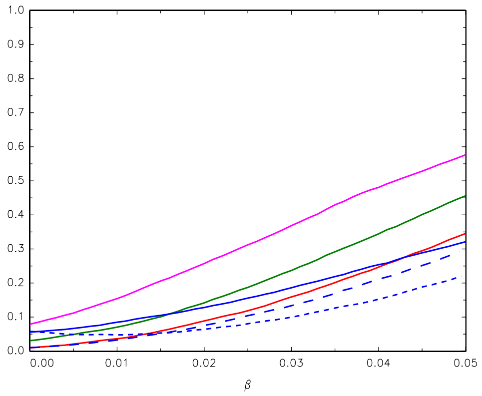
Figure S21. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.5$, $\delta = -0.5$; $\mathcal{T}_{hyb}$: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line, EMW: magenta solid line



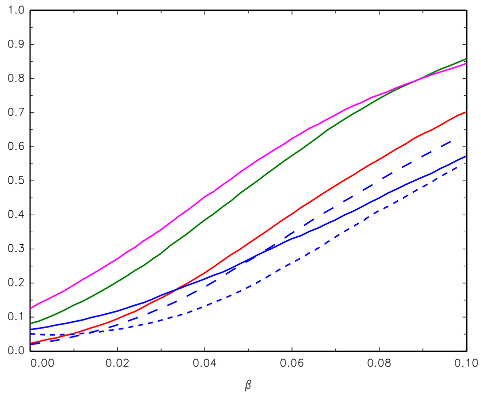
(a) $\phi = 1.025$



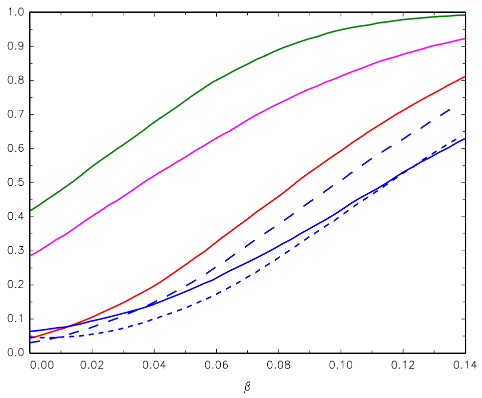
(b) $\phi = 1$



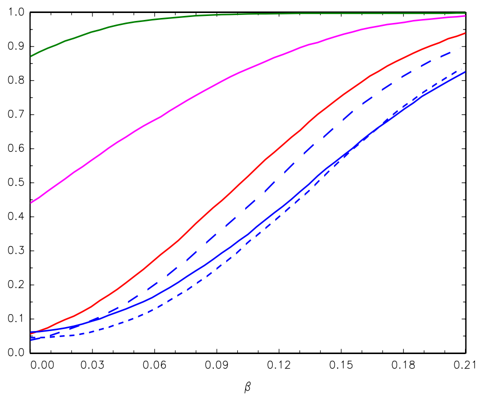
(c) $\phi = 0.975$



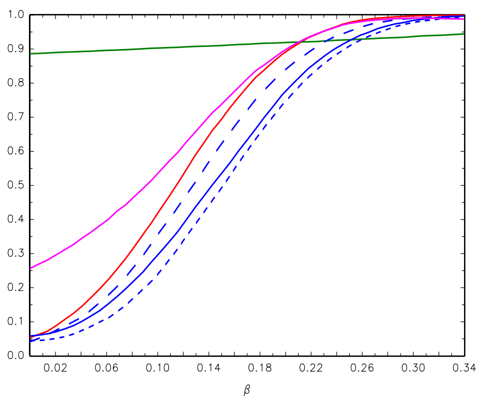
(d) $\phi = 0.95$



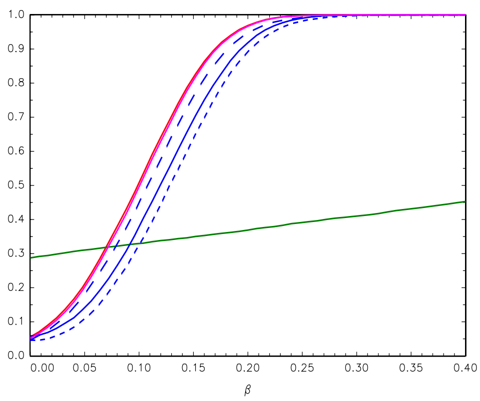
(e) $\phi = 0.875$



(f) $\phi = 0.75$



(g) $\phi = 0.5$



(h) $\phi = 0$

Figure S22. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.9$, $\delta = -0.5$; $\mathcal{T}_{hyb}$: ——, Q: ——, BD: ——, IVX_1: — —, IVX_2: - - -, EMW: ——