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

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

[a] School of Economics, University of Nottingham

[b] Essex Business School, University of Essex

October 2019

Abstract

Predictive regression methods are widely used to examine the predictability of (excess) stock returns by lagged financial variables characterised by unknown degrees of persistence and endogeneity. We develop new and easy to implement tests for predictability in these circumstances using regression t-ratios. The simplest possible test, optimal (under Gaussianity) for a weakly persistent and exogenous predictor, is based on the standard t-ratio from the OLS regression of returns on a constant and the lagged predictor. Where the predictor is endogenous, we show that the optimal, but infeasible, test for predictability is based on the t-ratio on the lagged predictor when augmenting the basic predictive regression above with the current period innovation driving the predictor. We propose a feasible version of this test, designed for the case where the predictor is an endogenous near-unit root process, using a GLS-based estimate of this innovation. We also discuss a variant of the standard t-ratio obtained from the predictive regression of OLS demeaned returns on the GLS demeaned lagged predictor. In the near-unit root case, the limiting null distributions of these three statistics depend on both the endogeneity correlation parameter and the local-to-unity parameter characterising the predictor. A feasible method for obtaining asymptotic critical values is discussed and response surfaces are provided. To develop procedures which display good size and power properties regardless of the degree of persistence of the predictor, we propose tests based on weighted combinations of the three t-ratios discussed above, where the weights are obtained using the p-values from a unit root test on the predictor. Using Monte Carlo methods we compare our preferred weighted test with the leading tests in the literature. These results suggest that, despite their simplicity, our weighted tests display very good finite sample size control and power across a range of persistence and endogeneity levels for the predictor, comparing very favourably with these extant tests. An empirical illustration using US stock returns is provided.

Keywords: predictive regression; persistence; endogeneity; weighted statistics.

JEL classification: C12, C22

*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 predictability tests from these regressions are of fundamental importance. These will depend on the time series properties of the predictors, in particular on their degree of persistence. Accordingly, several authors have argued that the findings from the studies cited above may be spurious. Nelson and Kim (1993) and Stambaugh (1999) show that highly persistent predictors lead to biased coefficient estimates from the predictive regressions if the innovations driving the predictors are correlated with returns (such that the predictive regressor is endogenous), as is argued to be the case for many of the popular macroeconomic and financial variables used as predictors; for example, the stock price is a component of both the return and the dividend yield. Goyal and Welch (2003) show that the persistence of dividend-based valuation ratios increased significantly over the typical sample periods used in empirical studies of predictability, and argue that, as a consequence, out-of-sample predictions using these variables are no better than those from a no-change strategy. Empirical evidence 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 predictors' innovations.

As a result, a number of predictability tests have been developed in the literature which are designed to be asymptotically valid when the predictor is strongly persistent and endogenous; see, among others, Cavanagh *et al.* (1995), Torous *et al.* (2004), CY, Kostakis *et al.* (2015), Breitung and Demetrescu (2015), Elliott *et al.* (2015) and Jansson and Moreira (2006). When such robust techniques are used the statistical evidence of predictability is considerably weaker and often disappears completely; see, among others, Ang and Bekaert (2007), Boudoukh *et al.* (2007), Welch and Goyal (2008) and Breitung and Demetrescu (2015). These methods are based on a formulation where a predictor, x_{t-1} say, is assumed to follow a first-order autoregression with coefficient $\phi = 1 - c/T$, where c is a finite non-negative constant and T is the sample size, which therefore approaches a random walk as T increases to infinity. In this case standard statistics, such as the associated regression t-ratio and likelihood ratio statistics, from the predictive regression have limiting distributions which depend on c and on the correlation between the innovations driving the predictor and returns; see, for example, Cavanagh *et al.* (1995).

Two methodological strands have developed in the literature to deliver asymptotically valid inference in this setting. The first remains based around the non-pivotal standard statistics from the predictive regression but uses methods of inference which account for the non-pivotal nature of these statistics. The leading example of this approach in the literature is CY who adapt the econometric methods developed in Cavanagh *et al.* (1995) based on inverting the non-pivotal limit distribution of the t-ratio and constructing, among other approaches, Bonferroni-type confidence intervals for the nuisance parameter c. The Q statistic of CY is widely regarded as the current state of the art methodology in the literature for testing the predictability of stock returns with highly persistent regressors.

In the second methodological strand, predictability tests are based on (asymptotically) pivotal statistics, obtained using alternative methods of estimating the predictive regression which take account of the properties of the regressor, rather than on the standard non-pivotal statistics. This has been achieved in two ways in the literature. The first, typified by the work of Toda and Yamamoto (1995), Dolado and Lutkepohl (1996), Bauer and Maynard (2012) and Breitung and Demetrescu (2015), uses variable addition whereby the predictive regression equation is augmented with additional (redundant) variables such that the coefficient of interest becomes associated with a stationary variable. Variable addition-based methods in general, however, have considerably lower power to reject against predictability than do the tests based on non-pivotal statistics discussed above. A second, more powerful approach, has been to use instrumental variable [IV] estimation of the predictive regression model. Phillips and Magdalinos (2009) and Kostakis *et al.* (2015) adopt this approach whereby each predictor in the predictive regression has an associated stochastic instrument formed by constructing a mildly integrated variable from the first differences of the predictor, what they term an extended IV or IVX instrument. The IVX instrument, by construction, has lower persistence than a local-to-unity variable and, as such, delivers an asymptotically pivotal predictability statistic. Breitung and Demetrescu (2015) discuss tests based on a variety of instruments fulfilling the same role as the IVX instrument in that, by construction, they are less persistent than a local-to-unity variable; they term these Type I instruments. They also discuss the use of Type 2 instruments, which may be either a non-stationary variable or a deterministic function of time, both satisfying the condition that it is independent of the associated predictor. They also discuss a testing procedure which combines both a Type 1 and a Type 2 instrument for each predictor.

A major practical drawback in implementing the first methodology relative to the second is that the method is invalid if the regressor is stationary or near-stationary; the theoretical validity of the method requires each predictor to be at least as persistent as a local-to-unity process. In contrast, the second approach is valid regardless of whether the predictors are local to unity or weakly dependent (stationary). In particular, the tests developed in Kostakis *et al.* (2015) and Breitung and Demetrescu (2015) based on a Type 1 instrument are designed for the case where the predictor is less persistent than a local-to-unity process, the leading case of interest being where it is stationary (weakly dependent). However, although still theoretically valid, their power is not as high as tests designed for local-to-unity predictors when the predictors are indeed strongly dependent. The combined instrument test of Breitung and Demetrescu (2015) is designed to mitigate against this such that, in large samples, it reverts to the Type 2 instrument when the predictor is local-to-unity but to the Type 1 instrument when the predictor is stationary. A very significant drawback with this approach, however, is that it can only be implemented as a two-tailed test and so when the direction of predictability from a given variable is known, it has considerably lower power than the one-sided tests.

In this paper we explore further how one can develop testing strategies which retain both good size properties and strong power profiles regardless of the degree of persistence of the predictor and which retain the flexibility that they can be implemented as either one-tailed or two-tailed tests. Our approach is focused on easy to implement tests using regression t-ratios. In particular, we propose a test based on a weighted combination of t-statistics, the weights depending on the persistence of the predictor via a function of the p-values from a standard Dickey-Fuller-type unit root test applied to the predictor. The constituent t-statistics are (i) a t-ratio on the lagged predictor when augmenting the basic predictive regression with a GLS-based estimate of the current period innovation driving the predictor (an infeasible version of this test using the actual current period innovation is optimal when the predictor is endogenous), and (ii) a t-ratio from a variant of the standard predictive regression where the OLS demeaned returns are regressed on the GLS demeaned lagged predictor. A further modification combines the weighted test with the standard t-ratio. In the near-unit root case, the proposed individual, weighted and modified weighted statistics have limiting null distributions which depend on both the endogeneity correlation parameter and the local-to-unity parameter characterising the predictor. We therefore propose a feasible method for obtaining asymptotic critical values and provide response surfaces for use in practice. We find that the newly proposed tests perform well in terms of finite sample size and power across a range of persistence levels for the predictor, and compare very favourably with extant tests, offering simple yet highly effective methods for predictability testing.

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 which we place on this data generating process [DGP]. These include the assumption that the innovations in the predictive regression model are independent and identically distributed (IID). In section 3, we present the three individual t-statistics that we employ and establish their asymptotic properties. Here we also outline our method for obtaining asymptotic critical values and use Monte Carlo simulation methods to compare the finite sample size and power properties of the resulting three t-tests. These simulation results provide motivation for the weighted statistics that we propose and evaluate in section 4. In section 5 we investigate the finite sample size and power properties of our proposed weighted tests and compare them with the leading tests in the literature, namely those due to Campbell and Yogo (2006), Kostakis *et al.* (2015) and Breitung and Demetrescu (2015). Section 6 contains an empirical illustration using monthly U.S. stock returns data. Section 7 concludes. Proofs are contained in a mathematical appendix.

2 The Predictive Regression Model

Let y_t denote the (excess) stock return in period t and let x_{t-1} denote a variable observed at time $t-1$ which is considered to be a putative predictor for y_t. The predictive regression model we consider is

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

where x_t is an observed process, specified according to the DGP

$$\begin{aligned} x_t &= \alpha_x + s_t, \qquad t = 1, ..., T \\ s_t &= \phi s_{t-1} + \epsilon_{xt}, \qquad t = 2, ..., T \end{aligned} \tag{2}$$

with s_1 a mean zero $O_p(1)$ random variable.

As discussed in the Introduction it is important for practical purposes to allow for the possibility of high persistence in the predictor variable x_t and to allow the shocks driving the predictor, ϵ_{xt} in (2), to be correlated with the unpredictable component of stock returns, ϵ_{yt} in (1). As regards the latter, we assume that the innovation vector $\epsilon_t := (\epsilon_{xt}, \epsilon_{yt})'$ is IID with finite fourth order moments and satisfying

$$\left[\begin{array}{c} \epsilon_{xt} \\ \epsilon_{yt} \end{array}\right] \sim IID\left(0, \left[\begin{array}{cc} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{array}\right]\right).$$

The assumption that ϵ_t is a vector IID process is made in order to simplify our presentation. We will later discuss in section 4 how the methods we propose can be modified to allow for weak dependence and/or heteroskedasticity in ϵ_t.

With respect to the degree of persistence in x_t, we allow ϕ in (2) to satisfy one of the following two assumptions:

Assumption S. Strongly persistent predictor: The autoregressive parameter ϕ in (2) is local-to-unity with $\phi := 1 - cT^{-1}$ where c is a fixed non-negative constant.

Assumption W. Weakly dependent predictor: The autoregressive parameter ϕ in (2) is fixed and bounded away from unity, $|\phi| < 1$.

Remark 1. Many 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. $\diamondsuit$

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), which do not require the practitioner to know which of Assumption S or Assumption W holds for ϕ in (2). The alternative hypothesis is that y_t is predictable by x_{t-1}, in which case $\beta \neq 0$. 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 new predictability tests we propose in this paper under local alternatives such that the slope parameter β in (1) is local-to-zero. The localisation 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 which of Assumption S and Assumption W holds. Where x_t is strongly persistent the appropriate local alternative is given by $H_{g,S} : \beta = gT^{-1}$, while for weakly dependent x_t, it is given by $H_{g,W} : \beta = gT^{-1/2}$, where in each case g is a finite constant. The different localisation rates reflect the fact that near-integration implies a much stronger signal from the predictor x_{t-1}.

The familiar Cholesky decomposition allows us to write the two components of ϵ_t in the form

$$\begin{aligned} \epsilon_{xt} &= \sigma_x e_{1t} \\ \epsilon_{yt} &= \sigma_y \left(\rho_{xy} e_{1t} + \sqrt{1-\rho_{xy}^2} e_{2t} \right) \end{aligned} \tag{3}$$

where $e_t := (e_{1t}, e_{2t})' \sim IID(0, I_2)$ and $\rho_{xy} := \sigma_{xy}/(\sigma_x \sigma_y)$ is the contemporaneous correlation between the innovations driving the predictor, ϵ_{xt}, and the unpredictable component of stock returns, ϵ_{yt}. Using this representation, we can then re-write the predictive regression in (1) as

$$y_t = \alpha_y + \beta x_{t-1} + \left(\frac{\sigma_y}{\sigma_x} \rho_{xy} \right) \epsilon_{xt} + \left(\sigma_y \sqrt{1-\rho_{xy}^2} \right) e_{2t}. \tag{4}$$

The representation in (4) is instructive, in that it demonstrates how a predictive regression featuring an endogenous predictor x_{t-1}, such as (1), can be re-written using ϵ_{xt} as an additional covariate in a form in which the predictor regressor, x_{t-1}, is strictly exogenous.

3 Predictability Tests

In what follows it is convenient to define a generically notated regression model:

$$y_t = \alpha + \beta x_{t-1} + \delta z_{xt} + v_t. \tag{5}$$

and consider the t-test associated with the OLS estimate of β in (5).

3.1 An Infeasible Test

If ϵ_{xt} was observed (abstracting from the constant α_x, this is equivalent to knowing ϕ), we could then perform a standard OLS regression in (5) with $z_{xt} = \epsilon_{xt}$, which is clearly a correct specification with respect to the DGP in (4). Denoting the corresponding infeasible t-statistic as t^{inf}, it is straightforward to show that t^{inf} has a standard normal limiting distribution under the null hypothesis H_0, irrespective of whether Assumption S or Assumption W holds. Moreover, under Gaussianity this would be an efficient test (among α_y, α_x invariant tests) whenever $\rho_{xy} \neq 0$. Note that including ϵ_{xt} as a regressor reduces the error variance from σ_y^2 in (1) to $\sigma_y^2(1-\rho_{xy}^2)$ in (4); that is, with knowledge of ϕ we can essentially subtract off the part of the innovation to returns that is correlated with the innovation to the predictor variable, thereby delivering a more powerful test. When $\rho_{xy} = 0$, t^{inf} remains asymptotically efficient as incorporation of the redundant regressor ϵ_{xt} has no effect in large samples.

3.2 Feasible Statistics

We now detail some simple feasible tests and informally discuss their properties.

3.2.1 A standard t-statistic

The standard t-statistic based on (1) is effectively based on an OLS regression that omits z_{xt} from (5). Denoting this statistic as t, it is straightforward to show that t has a standard normal limiting null distribution under Assumption W for any value of ρ_{xy}, and thus has the potential for nuisance parameter free inference in this world. With respect to the DGP in (4), t is based on a correctly specified regression when $\rho_{xy} = 0$, but when $\rho_{xy} \neq 0$, the regression omits a relevant regressor; while this does not affect the limiting null distribution, t will be inefficient relative to the infeasible test in the $\rho_{xy} \neq 0$ case. Under Assumption S, t has a standard normal limit null distribution provided $\rho_{xy} = 0$, in which case it is also efficient. When $\rho_{xy} \neq 0$, however, its limit null distribution depends on ρ_{xy} and c, and we would also expect the resulting test to be highly inefficient in this strongly persistent case, due to the lack of any proxy for ϵ_{xt} in the underlying regression.

3.2.2 A variant t-statistic

Before considering a statistic that includes a direct proxy for ϵ_{xt}, we also outline a variant of the standard t-statistic that might be considered more appropriate in the case of strongly persistent x_t. The standard t-statistic regression (1) is equivalent to regressing OLS demeaned y_t on OLS demeaned x_t, which implicitly uses $\bar{x}_{-1} := (T-1)^{-1}\sum_{t=2}^{T} x_{t-1} = O_p(T^{1/2})$ as an estimate of

α_x. A natural alternative to consider in the strongly persistent x_t context is a generalised least squares [GLS]-type demeaning for x_t, using $x_1 = O_p(1)$ as an estimator for α_x instead of $\bar{x}_{-1}$. We therefore consider a t-statistic associated with the OLS estimate of β in the model

$$(y_t - \bar{y}) = \beta(x_{t-1} - x_1) + v_t \tag{6}$$

which we denote by t'. As with t, the asymptotic null distribution of t' will depend on ρ_{xy} and c when x_t is strongly persistent. Under weakly dependent x_t, the null limit distribution of t' will depend on the (unknown) distribution of s_1, hence the statistic is only designed for use in the strongly persistent world (in contrast to t).

3.2.3 A t-statistic based on using a proxy measure for ϵ_{xt}

Given that ϵ_{xt} is unobservable, we might consider if it is possible to obtain a proxy measure for ϵ_{xt}. There are a number of ways in which this could be done, including the Bonferroni-based method advocated in CY. Here we consider an alternative approach based on including a covariate z_{xt} in (5) that acts as a direct proxy for ϵ_{xt} in (4). We will also focus our discussion for the present on the case of Assumption S where x_t is strongly persistent, as this is the most problematic case where the standard statistic t has a non-pivotal limiting distribution. An obvious approach to obtaining a proxy for ϵ_{xt} is to assume a particular value for the local-to-unity parameter c, say $\bar{c}$; we would then construct $z_{xt} = x_t - (1 - \bar{c}T^{-1})x_{t-1}$ (assuming $\alpha_x = 0$ for simplicity). If it happened to be the case that $\bar{c} = c$, then $z_{xt} = \epsilon_{xt}$ and we obtain the asymptotically standard normal and efficient test, $t^{\inf}$. However, when $\bar{c} \neq c$ the critical values for this test will depend on both ρ_{xy} and c, and it will no longer be an efficient test, with power being a (decreasing) function of the distance $|c - \bar{c}|$. This clearly poses a problem as c cannot be consistently estimated.

An obvious proxy for ϵ_{xt} is the OLS estimate, $\hat{\epsilon}_{xt}$ say, obtained from an OLS regression of Δx_t on a constant and x_{t-1}. However, setting $z_{xt} = \hat{\epsilon}_{xt}$ in (5) runs into the problem that z_{xt} is exact orthogonal to the predictive regressor x_{t-1}. The estimate of β from such a fitted model is then numerically identical to that which would be obtained if z_{xt} was omitted from (5). Moreover, the corresponding statistic is approximately $1/\sqrt{1 - \rho_{xy}^2}$ times the simple t-statistic, t, and so the inference drawn from such a test would essentially be identical to that from t, hence using the proxy regressor $\hat{\epsilon}_{xt}$ delivers no benefit.

An alternative method for obtaining a proxy for ϵ_{xt}, which takes account of a strongly persistent autoregressive structure in estimating the intercept term α_x, is to use a GLS-type estimator of α_x as in t'. To that end, consider regressing Δx_t on $(x_{t-1} - x_1)$, and denote the residuals from this fitted regression by $\tilde{\epsilon}_{xt}$.[1] Then consider setting $z_{xt} = \tilde{\epsilon}_{xt}$ in (5). In contradistinction to the

[1] Notice that this is based on the GLS-type estimator for α_x that would obtain on assuming that $c = \bar{c} = 0$. We also investigated choices of the GLS demeaning parameter, $\bar{c}$, other than $\bar{c} = 0$ but found very little difference in finite samples from the results given in this paper for $\bar{c} = 0$.

OLS-based proxy regressor $\hat{\epsilon}_{xt}$, the GLS-based proxy regressor $\tilde{\epsilon}_{xt}$ is not orthogonal to x_{t-1}. This lack of orthogonality raises the potential for $\tilde{\epsilon}_{xt}$ to act as a useful proxy for ϵ_{xt} in the strongly persistent case, and we consider the corresponding t-statistic, t^*, in what follows. As we will establish in section 3.3, the limiting null distribution of t^* depends on both ρ_{xy} and c in the case where x_t is strongly persistent (Assumption S), although this issue notwithstanding, we might anticipate that this procedure could deliver decent power performance due to the inclusion of a proxy for ϵ_{xt}. Under weak dependence of x_t (Assumption W), because s_t and s_1 are both of $O_p(1)$, the asymptotic distribution of $\tilde{\epsilon}_{xt}$, and therefore that of t^*, will depend on the distribution of s_1 (as with t').

3.3 Asymptotic Distributions of the Feasible Statistics

In Theorem 1 we now report the asymptotic distributions of the t, t' and t^* statistics under both the null and local alternatives for the case where x_t is strongly persistent. In Theorem 2 we subsequently present the corresponding limit for t in the case where x_t is weakly dependent.

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

$$
\begin{aligned}
t &\Rightarrow \frac{g\sigma_x}{\sigma_y}\sqrt{\int_0^1 \bar{W}_{1c}(r)^2 dr} + \frac{\int_0^1 \bar{W}_{1c}(r) d\left\{\rho_{xy} W_1(r) + \sqrt{1-\rho_{xy}^2} W_2(r)\right\}}{\sqrt{\int_0^1 \bar{W}_{1c}(r)^2 dr}} =: \mathcal{S}(g\sigma_x/\sigma_y, \rho_{xy}, c) \\
t' &\Rightarrow \frac{g\sigma_x}{\sigma_y}\frac{\int_0^1 \bar{W}_{1c}(r)^2 dr}{\sqrt{\int_0^1 W_{1c}(r)^2 dr}} + \frac{\int_0^1 \bar{W}_{1c}(r) d\left\{\rho_{xy} W_1(r) + \sqrt{1-\rho_{xy}^2} W_2(r)\right\}}{\sqrt{\int_0^1 W_{1c}(r)^2 dr}} =: \mathcal{S}'(g\sigma_x/\sigma_y, \rho_{xy}, c) \\
t^* &\Rightarrow \frac{g\sigma_x}{\sigma_y}\frac{\sqrt{\int_0^1 \bar{W}_{1c}(r)^2 dr}}{\sqrt{1-\rho_{xy}^2}} + \frac{\int_0^1 \bar{W}_{1c}(r) dW_2(r)}{\sqrt{\int_0^1 \bar{W}_{1c}(r)^2 dr}} + \frac{\rho_{xy}\sqrt{\int_0^1 \bar{W}_{1c}(r)^2 dr}\int_0^1 W_{1c}(r) dW_1(r)}{\sqrt{1-\rho_{xy}^2}\int_0^1 W_{1c}(r)^2 dr} \\
&\quad =: \mathcal{S}^*(g\sigma_x/\sigma_y, \rho_{xy}, c)
\end{aligned}
$$

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 2. The results in Theorem 1 highlight that for all three of the statistics considered the offset seen in their limiting distributions under the local alternative $H_{g,S}$, and hence their asymptotic local power, is a function of a common deterministic offset term, $g\sigma_x/\sigma_y$, in each case weighted by a different statistic-specific stochastic offset term. Under the null hypothesis, H_0, the asymptotic distributions of all three statistics are non-standard and depend on both ρ_{xy} and c. $\diamondsuit$

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

$$t \;\; \Rightarrow \;\; \frac{g\sigma_x}{\sigma_y} \frac{1}{\sqrt{1-\phi^2}} + N(0,1) =: \mathcal{W}(g\sigma_x/\sigma_y, \phi)$$

where $N(0,1)$ denotes a standard normal variate.

Remark 3. We do not present the corresponding limiting distributions for t' and t^* because, as discussed previously in sections 3.2.2 and 3.2.3, these depend on the distribution of s_1. The weighted test statistic, t^{wt}, which we will subsequently develop in section 4 is designed such that it never selects (i.e. gives zero weight to) t' or t^* in large samples under Assumption W and, hence, we will not need the limiting distribution of either t' or t^* to establish the limiting distribution of t^{wt} under Assumption W. ♢

Remark 4. Notice that, in contrast to what was seen under Assumption S, the asymptotic local power offset for t under Assumption W is purely non-stochastic being a (deterministic) function only of g, σ_x, σ_y and ϕ. Under H_0, t is seen to have a standard normal limiting distribution. ♢

3.4 Asymptotic Critical Values

Considering the case of strong dependence in Theorem 1 above, under H_0, the critical values of t, t',and t^* depend on the unknown parameters ρ_{xy} and c. At a practical level and as we will show below, ρ_{xy} can be consistently estimated so this dependence is easily dealt with. More important is the dependence on c, which cannot be dealt with as easily because c, unlike ρ_{xy}, is not consistently estimable. We therefore adopt a scheme for simulating critical values that will, by design, yield asymptotically conservative tests. We will illustrate this in the context of t and its null limit distribution $\mathcal{S}(0, \rho_{xy}, c)$, but the same approach can be used for t' and t^*, and their respective null limit distributions, $\mathcal{S}'(0, \rho_{xy}, c)$ and $\mathcal{S}^*(0, \rho_{xy}, c)$. For expository purposes we will focus attention on upper-tail critical values relevant for upper-tailed tests. The corresponding procedure for lower-tailed (or two-tailed) tests can be obtained in an entirely analogous fashion, or simply by using $-x_{t-1}$ rather than x_t as the predictor variable and carrying out upper-tailed tests as outlined below.

The steps to obtaining the conservative critical value are as follows:

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

2. At each value of c, compute the α-level upper-tail critical value, $cv_\alpha(\rho_{xy}, c)$ say.

3. Set the α-level critical value for t equal to $cv_\alpha(\rho_{xy}) := \max_{c\in[0,c_{\max}]} cv_\alpha(\rho_{xy}, c)$.

Using $cv_\alpha(\rho_{xy})$ will yield a correct α-level sized test when $c = \arg\max_{c\in[0,c_{\max}]} cv_\alpha(\rho_{xy}, c)$, and give a conservatively sized test for other values of c. We simulated critical values in this manner for all three tests, approximating the Brownian motion processes in the limiting functionals using $IIDN(0,1)$ random variates, and with the integrals approximated by normalised sums of 1,000 steps, with 20,000 replications. This was carried out for values of $\rho_{xy} \in [-0.950, -0.925, -0.900, ..., 0.900]$; we set $c_{\max} = 50$ and evaluated the critical values for $c \in [0, 2, 4, ..., 50]$ with $\alpha = 0.05$ and $\alpha = 0.10$. For most values of ρ_{xy}, $\arg\max_{c\in[0,c_{\max}]} cv_\alpha(\rho_{xy}, c)$ is obtained for c much smaller than $c_{\max}$; for example, with $\rho_{xy} = -0.9$, it is obtained at $c = 0$ for both values of α.

To automate selection of an appropriate critical value for a given value of ρ_{xy}, we calculated a response surface by regressing $cv_\alpha(z)$ on $[1, z, z^2, ..., z^8]$ with $z = \rho_{xy}$ for the 75 data points corresponding to the grid of values for ρ_{xy}. The response surface critical value is the fitted value from this regression. The response surface coefficient estimates can be found in Table 1, along with those for t' and t^*, denoted $cv'_\alpha(\rho_{xy})$ and $cv^*_\alpha(\rho_{xy})$, respectively. In practice, the response surface critical values can be calculated by substituting the unknown correlation parameter ρ_{xy} with a consistent estimate. To that end, we suggest using the estimator

$$\hat{\rho}_{xy} := \frac{\sum_{t=2}^{T} \hat{\epsilon}_{xt}\hat{\epsilon}_{yt}}{\sqrt{\sum_{t=2}^{T} \hat{\epsilon}_{xt}^2 \sum_{t=2}^{T} \hat{\epsilon}_{yt}^2}} \tag{7}$$

where $\hat{\epsilon}_{yt}$ are the OLS residuals from regressing y_t on a constant and x_{t-1}, and where it is recalled from section 3.2.3 that $\hat{\epsilon}_{xt}$ denote the OLS residuals from regressing Δx_t on a constant and x_{t-1}. It is straightforward to show that $\hat{\rho}_{xy}$ is a consistent estimator of ρ_{xy} under either Assumption S or Assumption W. In what follows, we denote tests based on comparison of t, t' and t^* with their asymptotically conservative critical values by t_C, t'_C and t^*_C, respectively.

3.5 Finite Sample Simulations

We now perform Monte Carlo simulation experiments to compare the finite sample size and power of the t_C, t'_C and t^*_C tests for $T = 200$ (other sample sizes gave qualitatively similar results). We generate (1)-(3) with $(e_{1t}, e_{2t})' \sim IIDN(0, I_2)$ and drawing s_1 as a standard normal variate, setting $\alpha_y = \alpha_x = 0$ without loss of generality. Although our theoretical results concerning t'_C and t^*_C are only established under Assumption S, in the simulations we deliberately blur the distinction between strong persistence and weak dependence of the predictor by setting $\phi = 1 - c/200$ and varying $c \in [0, 1, 2, ..., 200]$, such that x_t varies between a random walk and a white noise process. We conduct upper-tailed tests at the nominal 0.05 level, using critical values obtained from evaluating the estimated response surfaces from Table 1 at $\hat{\rho}_{xy}$ of (7), with

$\hat{\rho}_{xy}$ calculated separately for each Monte Carlo replication.[2] The simulation results reported in this section and in section 5 are all based on 20,000 Monte Carlo replications.

3.5.1 Finite Sample Size Properties

Figure 1 reports the simulated sizes of t_C, t'_C and t^*_C across c for different values of ρ_{xy}. By way of a benchmark for size, we also include sizes for a test that compares the statistic t with its asymptotic critical value in the weak dependence case, i.e. 1.645 from $N(0,1)$; this test is denoted t_N. Looking at panel (a) where $\rho_{xy} = -0.95$, we see that t_N is very badly over-sized for small c, while its empirical size gets closer to the nominal 0.05 level for large c. Its counterpart t_C, which uses the conservative response surface critical values, has near-correct empirical size with $c = 0$, but for $c > 0$ displays sizes which lie below the nominal level and increasingly so as c increases. Interestingly, although the response surface critical values are obtained under Assumption S, empirical sizes for $c = 50$ through to $c = 200$ seem to vary very little. The empirical sizes of t'_C and t^*_C appear roughly correct for small c, and while, like t_C, they fall below the nominal level for larger c, the rate at which this occurs is much less severe. Between t'_C and t^*_C, it is the former which holds on to size better for large values of c, which may have implications for the relative power of the two in this region. The results for $\rho_{xy} = -0.9$ in panel (b) are very similar to those for $\rho_{xy} = -0.95$. In panel (c) where $\rho_{xy} = -0.5$, while t_N remains over-sized, the other three tests tend to be less under-sized for large c, particularly t'_C. When $\rho_{xy} = 0$ in panel (d), all four tests display decent empirical size properties across c. As we would expect, the sizes for t_C and t_N are similar here. For the positive values of ρ_{xy} in panels (e) and (f), all the tests are under-sized for small c, while the sizes for t_N, t_C and t'_C all lie reasonably close to the nominal level for larger c. However, the empirical size of t^*_C diverges badly outside of the small c region, indicating that the asymptotic critical values obtained under the strong persistence assumption for c up to $c_{\max}$ are not appropriate here. Overall then, for negative correlation, our preference regarding finite sample size control would rest with t'_C for larger c, while we are ambivalent between t'_C and t^*_C for smaller c. For positive correlation, t^*_C would appear to be best avoided unless c is small.

3.5.2 Finite Sample Power Properties

Next, we plot simulated powers across g (≥ 0), with $\beta = g/200$, for different values of ρ_{xy} and c. Figure 2 considers $\rho_{xy} = -0.95$. First, we should note the powers of t_N are fairly meaningless for comparison purposes outside of panel (f) where $c = 200$, due to its over-size discussed in

[2]Lower-tailed tests can also be considered, and our simulation results regarding the size and power of t_C, t'_C, t^*_C, t^w_C and t^{wt}_C are also appropriate for lower-tail tests, provided the tests are either simply applied to $-x_{t-1}$ using the asymptotic critical values from section 3.4, or are applied to x_{t-1} but using the critical values $-cv_\alpha(-\rho_{xy})$, $-cv^*_\alpha(-\rho_{xy})$, $-cv^{**}_\alpha(-\rho_{xy})$, $-cv^w_\alpha(-\rho_{xy}, 2)$ and $-cv^{wt}_\alpha(-\rho_{xy}, 2, 0.01)$, respectively (with $\hat{\rho}_{xy}$ again replacing ρ_{xy} in the implementation of the tests).

section 3.5.1. Looking then at t_C, t'_C and t^*_C, we find that t_C and t^*_C are the most powerful tests when $c = 0$, with t'_C having somewhat lower power. For $c = 10$, t^*_C and t'_C now have similar power levels and have substantially greater power than t_C. The same is largely true for $c = 25$. For $c = 50$, 100 and 200, however, the power of t^*_C falls well behind that of t'_C. Note that for $c = 200$, t_N has correct size and, not surprisingly given the under-sizing seen in the other tests, the most power. On the basis of these power simulations alone, we are fairly ambivalent between t'_C and t^*_C representing the best performing test across small to moderate values of c, yet we hold a distinct preference for t'_C for the larger values of c. With $\rho_{xy} = -0.9$ (Figure 3), we draw similar conclusions, and again when $\rho_{xy} = -0.5$ (Figure 4). For $\rho_{xy} = 0$ (Figure 5), it becomes hard to draw any firm conclusions regarding the relative test rankings, since the power profiles are generally very similar across c (the only real exception to this is the lower power of t'_C when $c = 0$). When the correlation is positive (Figures 6 and 7), the tests again behave similarly when c is small (again excepting t'_C when $c = 0$), while for large c the t_C, t'_C and t_N tests have broadly similar power profiles, while the power for t^*_C is rendered uninformative because of its extreme over-size for large c.

4 Weighted Tests

Given the size and power simulation results reported in section 3.5, it would seem sensible to envisage a testing procedure based on suitably combining the statistics t' and t^*. Such a combination should involve both the statistics when c is of small to moderate magnitude, since neither of the t'_C and t^*_C tests are really dominant in terms of power across this region of c taken as a whole, and so it makes sense to permit both statistics to have the opportunity to provide evidence against the null. For larger c, the combination should revert to the t' statistic alone, because of the better power of the t'_C test, and the far superior size control when the correlation is positive.

An obvious way to achieve this is to consider a simple weighting scheme based on the p-value from a unit root test applied to x_t which has the usual property that the further is the dominant autoregressive root of x_t from unity, the closer is the p-value of the unit root test to zero, other things being equal. To that end, consider a generic unit root statistic, UR, which has the limiting distribution $\mathcal{S}^{UR}(c)$ under Assumption S. Denote by p^{UR} the p-value associated with UR using the limiting distribution for $c = 0$, $\mathcal{S}^{UR}(0)$. We require UR to be such that, under Assumption S, this p-value satisfies $p^{UR} = \int_{-\infty}^{UR} \mathcal{S}^{UR}(0) \Rightarrow \int_{-\infty}^{\mathcal{S}^{UR}(c)} \mathcal{S}^{UR}(0) =: p^{UR}(\mathcal{S}^{UR}(c))$. Notice that when $c = 0$, $p^{UR} \Rightarrow p^{UR}(\mathcal{S}^{UR}(0)) = U(0,1)$. Based on p^{UR} we can then define a weighted statistic, denoted t^w, as

$$t^w := (p^{UR})^\lambda t^* + \{1 - (p^{UR})^\lambda\} t' \tag{8}$$

where λ is a positive constant. So, for small c, p^{UR} will be non-zero and t^w will combine inference from both t^* and t'. For larger c, p^{UR} will tend to be smaller (with x_t appearing less persistent) and so the majority of the weighting in t^w will shift to t'. For very large c, p^{UR} is essentially zero, at which point t^w coincides with t'.

Theorem 3. *Under the conditions of Theorem 1,*

$$\begin{aligned} t^w &\Rightarrow (p^{UR}(c))^\lambda \mathcal{S}^*(g\sigma_x/\sigma_y, \rho_{xy}, c) + \{1-(p^{UR}(c))^\lambda\}\mathcal{S}'(g\sigma_x/\sigma_y, \rho_{xy}, c) \\ &=: \mathcal{S}^w_{UR}(g\sigma_x/\sigma_y, \rho_{xy}, c, \lambda). \end{aligned}$$

In addition to t^w, we also consider a further modification designed to potentially improve power in cases of low persistence. Consider the case where the x_t process is white noise; here, the t_N test is valid, and Figures 2-7 highlight an attractive power profile for this test - see the $c = 200$ results. (For negative ρ_{xy}, the power results for the t_C test in the figures are adversely impacted by the fact that the statistic t is being compared with conservative critical values, which are dominated by the behaviour of t at $c = 0$.) It makes sense, therefore, to consider whether additional power can be harnessed from integrating the statistic t into the weighted test when c is very large. A refinement of t^w can therefore be considered where a switch into t occurs for very large c. Specifically, we consider a statistic of the form

$$t^{wt} := t^w \mathbb{I}(p^{UR} > \tau) + t\mathbb{I}(p^{UR} \leq \tau)$$

where $\mathbb{I}(.)$ is the indicator function and $\tau > 0$. Here t^{wt} is the same as t^w for values of p^{UR} above a user-specified cut-off value τ; otherwise it is simply t. The motivation here is that if τ is chosen suitably small, then for small c, t^{wt} will typically be t^w rather than t, so that t cannot inflate the critical values of t^{wt} very much in this region. Then, for large c, when t^{wt} is typically t, t^{wt} will be using less conservative critical values than those required to control the size of t_C, thereby improving power in this region.

Theorem 4. *Under the conditions of Theorem 1,*

$$\begin{aligned} t^{wt} &\Rightarrow \mathcal{S}^w(g\sigma_x/\sigma_y, \rho_{xy}, c, \lambda)\mathbb{I}(p^{UR}(c) > \tau) + \mathcal{S}(g\sigma_x/\sigma_y, \rho_{xy}, c)\mathbb{I}(p^{UR}(c) \leq \tau) \\ &=: \mathcal{S}^{wt}_{UR}(g\sigma_x/\sigma_y, \rho_{xy}, c, \lambda, \tau). \end{aligned}$$

Remark 5. It can be seen from Theorem 3 and Theorem 4 that the asymptotic distributions of t^w and t^{wt} depend in general on ρ_{xy}, c, the choice of λ, and on the specific unit root test statistic, UR, used in defining the weights. Additionally, the limiting distribution of t^{wt} depends on the choice of τ. $\diamondsuit$

Remark 6. Under the conditions of Theorem 2, so that Assumption W holds, it holds trivially that t^{wt} and t are asymptotically equivalent, by virtue of that fact that the value of τ used in (8) is a positive constant. Consequently, $t^{wt} \Rightarrow \mathcal{W}(g\sigma_x/\sigma_y, \phi)$. $\diamondsuit$

We next consider how to obtain simulated critical values that will yield asymptotically conservative tests based on t^w and t^{wt}. These critical values will depend on the specific choice made for the unit root statistic, UR. In what follows we report results for the case where UR is chosen to be the familiar local-GLS detrended Dickey-Fuller unit root test statistic of Elliott *et al.* (1996) for the choice of $\bar{c} = 0$; that is, the t-statistic for $\phi = 0$ associated with the OLS estimate of ϕ in the model

$$\Delta x_t = \phi(x_{t-1} - x_1) + v_t. \tag{9}$$

We will denote this t-statistic by DF in what follows so that under this choice $UR = DF$. Well known results show that, under Assumption S, $DF \Rightarrow \frac{\int_0^1 W_{1c}(r)dW_{1c}(r)}{\sqrt{\int_0^1 W_{1c}(r)^2 dr}} =: \mathcal{S}^{DF}(c)$, and the associated p-value, p^{DF}, satisfies the generic requirements for UR stated above. We have chosen DF because it is both simple to compute and we found that it gave rise to weighted tests with superior finite sample properties to other commonly used unit root tests, including the corresponding OLS detrended Dickey-Fuller unit root test. In what follows for $UR = DF$ we set $\lambda = 2$ and $\tau = 0.01$; we denote the associated asymptotic null distributions of t^w and t^{wt} as $\mathcal{S}^w_{DF}(\rho_{xy}, c, 2)$ and $\mathcal{S}^{wt}_{DF}(\rho_{xy}, c, 2, 0.01)$, respectively.[3]

Given the foregoing choice for UR, the next step is to simulate the limit Dickey-Fuller p-values. This could be done using a suitable bootstrap method, but here we outline a self-contained simulation-based method. To that end, we simulated $\mathcal{S}^{DF}(0)$ using $IIDN(0,1)$ random variates, and with the integrals approximated by normalized sums of 1000 steps, with 20000 replications, and then calculated the numerical approximation to $p^{DF}(x)$ for $x \in [-5, -4.95, -4.90, ..., 3]$. To automate selection of an appropriate asymptotic p-value for a given value of x, we once again calculated a response surface by regressing $p^{DF}(x)$ on $[1, z, z^2, ..., z^{11}]$ with $z = 1/(1 + e^{-x})$ (161 data points), the function z being a convenient choice given that we are approximating a cumulative density function. The response surface p-value is the fitted value from this regression, and the response surface coefficient estimates can be found in Table 2, denoted $p^{DF}(x)$. In practice, the response surface p-value can be calculated using $x = DF$. Calculation of conservative asymptotic critical values for the t^w and t^{wt} statistics is then carried out in exactly the same manner as in calculating those for t, t' and t^* in section 3.4 above, but based on $\mathcal{S}^w_{DF}(\rho_{xy}, c, 2)$ and $\mathcal{S}^{wt}_{DF}(\rho_{xy}, c, 2, 0.01)$. The response surface coefficient estimates for the ρ_{xy}-dependent critical values can be found in Table 2, denoted $cv^w_\alpha(\rho_{xy}, 2)$ and $cv^{wt}_\alpha(\rho_{xy}, 2, 0.01)$, respectively. We denote the tests based on comparison of t^w and t^{wt} with their asymptotically conservative critical values by t^w_C and t^{wt}_C, respectively.

Remark 7. Thus far we have assumed that ϵ_t is a vector IID process. It is straightforward to

[3]The choices of $\lambda = 2$ and $\tau = 0.01$ do not arise from analytical considerations; they were made on the basis of experimentation and finite sample size and power simulations. Other values can be used, but, somewhat inevitably, a trade-off will arise with respect to the performance of the tests across different parameter settings in the DGP.

modify the testing procedures discussed in this paper to allow for weak (stationary) autocorrelation and/or certain forms of heteroskedasticity in ϵ_t. For the former, the limiting null results given to date will continue to hold when the innovation process ϵ_{xt} in (2) is replaced by a stationary autoregressive process of the form $v_{xt} = \psi_1 v_{x,t-1} + \psi_2 v_{x,t-2} + ... + \psi_p v_{x,t-p} + \epsilon_{xt}$, provided the regressions involving x_t underlying the computation of the residuals $\hat{\epsilon}_{xt}$ and $\tilde{\epsilon}_{xt}$ defined in section 3.2.3 and the unit root regression in (9) from which the statistic DF is obtained are augmented with the additional regressors, $\Delta x_{t-1}, \Delta x_{t-2}, ..., \Delta x_{t-p}$. In practice, p will be unknown and the number of lagged difference regressors to include can be determined using any of the standard model selection criteria, for example BIC. One can also weaken the IID assumption to allow for conditional heteroskedasticity in the innovations without altering the asymptotic null properties of our proposed test statistics provided, as in Kostakis *et al.* (2015, p.1516), one uses White standard errors rather than OLS standard errors in constructing each of the t, t' and t^* statistics defined in section 3. Indeed, when basing the t, t' and t^* statistics on White standard errors, unconditional heteroskedasticity in ϵ_{yt} of the form considered in, for example, Assumption 1 of Breitung and Demetrescu (2015,p.359) can also be permitted. Finally, it should be noted that the presence of either stationary autocorrelation or heteroskedasticity will alter the limiting local power properties of the t, t', t^*, t^w and t^{wt} statistics (and the modifications of these discussed above) as they also do other tests in the extant literature including the tests of CY, Kostakis *et al.* (2015) and Breitung and Demetrescu (2015). $\diamondsuit$

In the next section we examine the finite sample size and power properties of our proposed weighted tests and compare them with the leading tests in the literature, namely those of Campbell and Yogo (2006), Kostakis *et al.* (2015) and Breitung and Demetrescu (2015).

5 Finite Sample Simulations

Figures 1 and 2-7 also show the simulated finite sample sizes and powers of t_C^w and t_C^{wt} for the same simulation design settings as were used in section 3.5, and again based on upper-tail tests conducted at the nominal 0.05 level using the estimated response surfaces from Table 1, again evaluated at $\hat{\rho}_{xy}$ of (7). Both $\hat{\rho}_{xy}$ and the Dickey-Fuller p-value, p^{DF} (obtained from the estimated response surface in Table 1), were calculated separately in each Monte Carlo replication. Examining Figure 1, and considering first t_C^w, we see a striking similarity between its size behaviour and that of t'_C. Only for $\rho_{xy} = 0.9$ in panel (f) do their sizes appear to differ in any way, with t_C^w being comparatively more under-sized across c. A notable feature is that for the positive values of ρ_{xy}, t_C^w does not exhibit the large over-sizing outside of small c associated with t_C^*. The reason underlying this is that for the larger c values considered, p^{DF} will tend to be small, with the result that t^* will generally receive a low weight in t^w. As regards t_C^{wt}, it has accurate size across c for all the negative values of ρ_{xy} considered, establishing it quite convincingly as the most reliably sized test in those cases. Elsewhere, it has a tendency to some under-sizing for small c, although since this is generally a feature of all the other tests, t_C^{wt} is still very competitive. Comparing t_C^w and t_C^{wt}, the main feature is that for the larger values of c, t_C^{wt} is notably less under-sized, reflecting its switch into t alone when p^{DF} is small. Interestingly, t_C^w and t_C^{wt} are the only approximately correctly sized tests when $\rho_{xy} = 0.9$ and $c = 0$.

Figure 2 also graphs the powers of t_C^w and t_C^{wt} for $\rho_{xy} = -0.95$. When $c = 0$ (ignoring t_N due to its over-size) we see that t_C^w and t_C^{wt} emerge as comfortably the most powerful tests, with both well exceeding the power of the third placed tests t_C and t^*, and there is little to choose between t_C^w and t_C^{wt} on power here. For $c = 10$ and $c = 25$, t'_C, t^*_C, t_C^w and t_C^{wt} show broadly similar power profiles. However, once $c = 50$ it is t_C^{wt} that emerges as the most powerful test, with t_C^w tying for second place with t'_C. These characteristics arise because t^w is now placing most weight on t', while t^{wt} is now very regularly switching into t. The same rankings are seen for $c = 100$ and 200, and here we also see that the power advantage that t_C^{wt} holds over t_C^w increases with c. Notice that when $c = 200$, t_C^{wt} has power approaching the levels associated with t_N (which is correctly sized here). These same comments regarding power ranking also hold when $\rho_{xy} = -0.9$ (Figure 3) and where $\rho_{xy} = -0.5$ (Figure 4). When $\rho_{xy} = 0$ (Figure 5) neither t_C^w or t_C^{wt} dominates the best of the other tests, but the power differences involved are really very modest. For $\rho_{xy} = 0.5$ (Figure 6), it is arguably t_C^{wt} that is the best performing test overall, albeit by a small margin. For the larger c values, we see that t_C^{wt} again outperforms t_C^w. Finally, for $\rho_{xy} = 0.9$ (Figure 7) when $c = 0$, t_C^w and t_C^{wt} are easily the best performing tests, by virtue of the other tests having very low empirical size here, as discussed above. For large c, the dominance of t_C^{wt} over t_C^w is quite evident.

On the basis of all of our simulation results for the tests considered, it seems reasonable to conclude that t_C^{wt} offers the best overall combination of finite sample size control and provision of power. We therefore next compare t_C^{wt} with the leading predictability tests currently available in the literature. Specifically, the tests we compare t_C^{wt} against are CY's Q test (one-sided, upper tail), the instrumental variable test of Breitung and Demetrescu (2015) using sine and fractional instruments, denoted BD (recall that this test can only be run as a two-sided test), and the IVX test of Kostakis *et al.* (2015), with IVX_1 and IVX_2 denoting one-sided (upper tail) and two-sided tests, respectively. All tests were implemented using settings appropriate for IID errors (we fixed the AR lag to one in Q, and used short run variance estimators in IVX, setting $M_n = 0$ in the notation of Kostakis *et al.*, 2015). As discussed in Kostakis *et al.* (2015, p.1514) the IVX instrument does not need to be demeaned because the slope estimator in the predictive regression is invariant to whether the instrument is demeaned or not. In calculating the IVX_1 and IVX_2 tests we implemented the finite-sample correction factor outlined in Kostakis *et al.* (2015, p.1516). Figures 8 and 9-14 graph the simulated finite sample empirical sizes and powers of these tests together with those of t_C^{wt} for the same DGP settings as were used in section 3.5. All tests were run at the nominal 0.05 (asymptotic) significance level and the plots for t_C^{wt} are reproduced from Figures 1 and 2-7 for convenience.

The standout feature from Figure 8 is that CY's Q test is very badly over-sized, for any value of $\rho_{xy} \neq 0$, unless c is small. The invalidity of the Q test for weakly stationary x_t is clearly reflected in Figure 8 as c becomes large. Another prominent feature is that the lower-tailed IVX_1 test is badly over-sized for negative ρ_{xy} and under-sized to a similar degree for positive ρ_{xy} when c is small. Of all of the tests considered, BD arguably appears to offer the most precise finite sample size control overall, followed by t_C^{wt} and IVX_2. Among the one-sided tests it is fair to conclude that t_C^{wt} offers the best size performance, in particular that it avoids the issue of over-sizing seen with the other one-sided tests.

Figure 9 presents the powers for $\rho_{xy} = -0.95$. For $c = 0$ (ignoring IVX_1 due to its significant over-size discussed above) we see that t_C^{wt} is generally the most powerful test, outperforming Q for all but the larger values of g considered and easily dominating both IVX_2 and BD. There appears to be little to choose between t_C^{wt}, Q and IVX_2 for $c = 10$ and $c = 25$, BD having comparatively very low power here. IVX_1 and t_C^{wt} are the best performing tests for the larger values of c considered (IVX_1 no longer being over-sized), and here there is relatively little to separate them. It is also interesting to note that Q is both over-sized *and* has poor power here. The results for $\rho_{xy} = -0.9$ in Figure 10 are very similar to those for $\rho_{xy} = -0.95$, and so similar comments apply. When $\rho_{xy} = -0.5$ (Figure 11), Q emerges as the most powerful test when $c = 0$ (IVX_1 is still over-sized here for small c) but is only marginally more powerful than t_C^{wt}. For moderate and large values of c, IVX_1 and t_C^{wt} provide the highest powers and are similar to each other. Again, Q has poor properties for large c. For $\rho_{xy} = 0$ (Figure 12) there is generally little to choose between t_C^{wt}, Q and IVX_1. When $\rho_{xy} = 0.5$ (Figure 13), arguably it is t_C^{wt} that has the best power performance overall. A similar claim could legitimately be made when $\rho_{xy} = 0.9$ (Figure 14), particularly given the performance of t_C^{wt} for $c = 0$.

Based on these simulation results, we conclude that t_C^{wt} 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 single 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, t_C^{wt} 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.

6 An Empirical Illustration

To illustrate how our proposed tests might behave in practice, we apply them to the monthly U.S. annual equity series analysed in Welch and Goyal (2008), using updated data for the period

1970:1-2017:12 ($T = 576$) which is available at http://www.hec.unil.ch/agoyal/. Our dependent variable y_t is the 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). Our test procedures t_C, t'_C, t^*_C, t^w_C and t^{wt}_C, employing $UR = DF$, $\lambda = 2$ and $\tau = 0.01$, are applied to these series with the number of lagged difference terms in Δx_t added to the underlying OLS regressions for $\hat{\epsilon}_{xt}$ and $\tilde{\epsilon}_{xt}$ and DF determined using BIC selection starting from a maximum value of $p_{\max} = 12$. Critical values for these tests were determined using the same method as detailed previously in connection with the finite sample simulation results. We also apply BD, IVX_1 and IVX_2 and Q (the latter being one-sided, applied at the asymptotic 0.05 level). The IVX statistic was calculated using long run variance estimators with $M_n = \lfloor T^{1/3} \rfloor$ in the notation of Kostakis *et al.* (2015), and Q was implemented with BIC lag selection from a maximum AR lag order of $p_{\max} + 1$.

Table 2 shows the results for those of the thirteen series where at least one of the nine tests considered yields a rejection at the 0.10 level. There are five such cases: the dividend-price ratio (denoted d/p), dividend yield (d/y), default return spread (dfr), inflation rate (inf) and the stock variance ($svar$). Here, we also report the values of $\hat{\rho}_{xy}$ and p^{DF}. We observe from Table 2 that for the strongly persistent predictors d/p and d/y, persistence being measured by p^{DF}, evidence of predictability at the 0.10 level is found only by (at least one of) our new tests. Indeed, for d/p, where in addition to strong persistence the value of $\hat{\rho}_{xy}$ indicates a very high degree of negative correlation, only our preferred weighted test, t^{wt}_C, delivers a rejection. This fits nicely into the sort of environment where our simulations showed that t^{wt}_C will be more powerful than t_C, t'_C and t^*_C. The remaining predictors dfr, inf and $svar$ do not appear to be strongly persistent since in each case p^{DF} is close to zero. Here we see that t^w is placing virtually all its weight on t', while in the case of dfr and $svar$, this low persistence further prompts the t^{wt} statistic to switch into t. For dfr, this switch turns a 0.10-level non-rejection by t^w_C into a 0.05 level rejection by t^{wt}_C. This again accords well with our simulation evidence which showed t^{wt}_C to be more powerful than t^w_C under weak persistence. All of the tests considered show 0.05 level rejections for svar. Notably, Q demonstrates 0.05 level rejections for all three of the non-persistent predictors - unlike any of the other tests which manage rejections for at most two of these predictors (i.e. t_C, t^*_C, and t^{wt}_C). We would speculate that the tendency of Q to be over-sized in the presence of low-persistence predictors, as was observed in the simulation results in section 5, may partly explain these findings. So, focussing on the outcomes of t^{wt}_C in the context of the new procedures introduced in this paper, we find that it uncovers at least as much evidence for predictability in these series as any of its comparator tests, notwithstanding the more questionable evidence arising from the Q test.

7 Conclusions

In this paper we have developed new and easy to implement tests for predictability based on simple to compute regression t-ratios. In particular, together with the standard t-ratio from the OLS regression of returns on a constant and a lagged predictor, we have discussed: (i) a t-ratio on the lagged predictor when augmenting the basic predictive regression with a GLS-based estimate of the current period innovation driving the predictor (an infeasible version of this test using the actual current period innovation is optimal when the predictor is endogenous), and (ii) a t-ratio from a variant of the standard predictive regression where the OLS demeaned returns are regressed on the GLS demeaned lagged predictor. We have proposed a feasible method for obtaining (conservative) asymptotic critical values for the tests based on each of these statistics and associated response surfaces have been provided. Finite sample simulations have been used to assess the size and power of these three tests across different degrees of predictor persistence and degrees of innovation endogeneity. Subsequently, to develop procedures which display good size and power properties regardless of the degree of persistence and endogeneity, we have also proposed tests based on weighted combinations of the t-ratios in (i) and (ii), the weights obtained according to the p-values from a standard Dickey-Fuller-type unit root test on the predictor and designed to place more weight on the test of (ii) in the low persistence environment. A further modification combines the weighted test with the standard t-ratio which reverts to the latter under very low persistence. The modification in particular yields a test procedure that compares very favourably with the leading tests for predictability in the literature, offering arguably the best trade-off of in terms of finite sample size and power properties overall across a broad diversity of persistence and endogeneity settings. In an empirical application to a well-known data set of US stock returns and putative predictors, we uncover evidence of predictability using our new tests for a number of predictors with significantly differing persistence and endogeneity characteristics that is not matched by competitor procedures with known reliable size.

References

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

Ang, A. and G. Bekaert (2007). Stock return predictability: Is it there? *Review of Financial Studies* 20, 651–707.

Bauer, D. and A. Maynard (2012). Persistence-robust surplus-lag Granger causality testing. *Journal of Econometrics* 169, 293–300.

Bickel, P. and M. Wichura (1971). Convergence criteria for multiparameter stochastic processes and some applications. *The Annals of Mathematical Statistics* 42, 1656–1670.

Boudoukh, J. R., M. Michaely, P. Richardson and M. R. Roberts (2007). On the importance of measuring payout yield: Implications for empirical asset pricing. *Journal of Finance* 62, 877–915.

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.

Dolado, J. J. and H. Lütkepohl (1996). Making Wald tests work for cointegrated VAR systems. *Econometric Reviews* 15, 369–386

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.

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.

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.

Goyal, A. and I. Welch (2003). Predicting the equity premium with dividend ratios. *Management Science* 49, 639–654.

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.

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

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.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics* 54, 375–421.

Toda, H. Y. and T. Yamamoto (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics* 66, 225–250.

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

A Appendix

Proof of Theorem 1: In what follows we may set $\alpha_y = \alpha_x = 0$; this is without loss of generality in view of the fact that all of the test statistics we consider are exact invariant to these parameters. We adopt the generic notation of (5), (6) and (9), denoting OLS estimators of the parameters of these models as $\hat{\alpha}$, $\hat{\beta}$, $\hat{\delta}$ and $\hat{\phi}$, as appropriate. Summations are taken over $t = 2, ..., T$, and integrals over $[0, 1]$, unless otherwise stated.

We will make use the following weak convergence results:

$$T^{-1/2}\sum_{t=1}^{\lfloor Tr \rfloor} e_{1t} \Rightarrow W_1(r), \quad T^{-1/2}\sum_{t=1}^{\lfloor Tr \rfloor} e_{2t} \Rightarrow W_2(r)$$

where $[W_1(r),\, W_2(r)]'$ is a bivariate standard Brownian Motion process. Then we can also write

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{\lfloor Tr \rfloor} \epsilon_{xt} &\Rightarrow \sigma_x W_1(r) \\ T^{-1/2}\sum_{t=1}^{\lfloor Tr \rfloor} \epsilon_{yt} &\Rightarrow \sigma_y \left\{\rho_{xy} W_1(r) + \sqrt{1-\rho_{xy}^2}\, W_2(r)\right\} \end{aligned}$$

and, under Assumption S, $T^{-1/2} x_{\lfloor Tr \rfloor} \Rightarrow \sigma_x W_{1c}(r) = \sigma_x \int_0^r e^{-(r-s)c} dW_1(s)$.

(i) We can write

$$t = \frac{T^{-1}\sum(x_{t-1} - \bar{x}_{-1}) y_t}{\sqrt{\hat{\sigma}_v^2 T^{-2} \sum(x_{t-1} - \bar{x}_{-1})^2}}$$

where $\bar{x}_{-1} := (T-1)^{-1}\sum x_{t-1}$, $\hat{\sigma}_v^2 := T^{-1}\sum \hat{v}_t^2$ and $\hat{v}_t := y_t - \hat{\alpha} - \hat{\beta} x_{t-1}$. Standard results, and the fact that $\hat{\sigma}_v^2 \xrightarrow{p} \sigma_y^2$, establish that

$$t \Rightarrow \frac{g\sigma_x}{\sigma_y}\sqrt{\int \bar{W}_{1c}(r)^2 dr} + \frac{\int \bar{W}_{1c}(r) d\left\{\rho_{xy} W_1(r) + \sqrt{1-\rho_{xy}^2}\, W_2(r)\right\}}{\sqrt{\int \bar{W}_{1c}(r)^2 dr}}$$

(ii) Here, we have

$$t' = \frac{T^{-1}\sum(x_{t-1} - x_1)(y_t - \bar{y})}{\sqrt{\hat{\sigma}_v^2 T^{-2}\sum(x_{t-1} - x_1)^2}}$$

where $\hat{\sigma}_v^2 := T^{-1}\sum \hat{v}_t^2$ and $\hat{v}_t := y_t - \hat{\alpha} - \hat{\beta} x_{t-1}$. Then $\hat{\sigma}_v^2 \xrightarrow{p} \sigma_y^2$, $T^{-2}\sum(x_{t-1} - x_1)^2 \Rightarrow$

$\sigma_x^2 \int W_{1c}(r)^2 dr$, and

$$\begin{aligned}
T^{-1}\sum(x_{t-1}-x_1)(y_t-\bar{y}) &= T^{-1}\sum(x_{t-1}-\bar{x}_{-1})(y_t-\bar{y}) \\
&= T^{-1}\sum(x_{t-1}-\bar{x}_{-1})y_t \\
&\Rightarrow g\sigma_x^2 \int \bar{W}_{1c}(r)^2 dr + \sigma_x\sigma_y \int \bar{W}_{1c}(r) d\left\{\rho_{xy}W_1(r)+\sqrt{1-\rho_{xy}^2}W_2(r)\right\}
\end{aligned}$$

Hence

$$t' \Rightarrow \frac{g\sigma_x}{\sigma_y}\frac{\int \bar{W}_{1c}(r)^2 dr}{\sqrt{\int W_{1c}(r)^2 dr}} + \frac{\int \bar{W}_{1c}(r) d\left\{\rho_{xy}W_1(r)+\sqrt{1-\rho_{xy}^2}W_2(r)\right\}}{\sqrt{\int W_{1c}(r)^2 dr}}.$$

(iii) We will use the following easily shown results. First, that $T^{-1}\sum(y_t-\bar{y})^2 \xrightarrow{p} \sigma_y^2$ and $T^{-1}\sum\tilde{\epsilon}_{xt}^2 \xrightarrow{p} \sigma_x^2$ with $\tilde{\epsilon}_{xt} := \Delta x_t - \hat{\phi}(x_{t-1}-x_1)$. Next,

$$T\hat{\phi} \;\Rightarrow\; \frac{\int W_{1c}(r)dW_{1c}(r)}{\int W_{1c}(r)^2 dr} = c + \frac{\int W_{1c}(r)dW_1(r)}{\int W_{1c}(r)^2 dr} := L(c)$$

$$\begin{aligned}
T^{-1}\sum(x_{t-1}-\bar{x}_{-1})\tilde{\epsilon}_{xt} &= T^{-1}\sum(x_{t-1}-\bar{x}_{-1})\{\Delta x_t - \hat{\phi}(x_{t-1}-x_1)\} \\
&= T^{-1}\sum(x_{t-1}-\bar{x}_{-1})\Delta x_t - T\hat{\phi}T^{-2}\sum(x_{t-1}-\bar{x}_{-1})x_{t-1} \\
&\Rightarrow \sigma_x^2\int \bar{W}_{1c}(r)dW_{1c}(r) - \sigma_x^2 L(c)\int \bar{W}_{1c}(r)^2 dr
\end{aligned}$$

$$T^{-1}\sum(x_{t-1}-\bar{x}_{-1})y_t \Rightarrow g\sigma_x\int \bar{W}_{1c}(r)^2 dr + \sigma_x\sigma_y\int \bar{W}_{1c}(r) d\left\{\rho_{xy}W_1(r)+\sqrt{1-\rho_{xy}^2}W_2(r)\right\}$$

$$\begin{aligned}
T^{-1}\sum\tilde{\epsilon}_{xt}(y_t-\bar{y}) &= T^{-1}\sum\tilde{\epsilon}_{xt}(gT^{-1}(x_{t-1}-\bar{x}_{-1}) + \epsilon_{yt}-\bar{\epsilon}_y) \\
&= gT^{-2}\sum(x_{t-1}-\bar{x}_{-1})\tilde{\epsilon}_{xt} + T^{-1}\sum\tilde{\epsilon}_{xt}(\epsilon_{yt}-\bar{\epsilon}_y) \\
&= T^{-1}\sum\{\Delta(x_t-x_1)-\hat{\phi}(x_{t-1}-x_1)\}(\epsilon_{yt}-\bar{\epsilon}_y) + o_p(1) \\
&= T^{-1}\sum(\epsilon_{yt}-\bar{\epsilon}_y)\Delta x_t + o_p(1) \Rightarrow \sigma_{xy}
\end{aligned}$$

$$\begin{aligned}
T^{-1}\sum\hat{v}_t^2 &= T^{-1}\sum\{(y_t-\bar{y}) - \hat{\beta}'(x_{t-1}-\bar{x}_{-1}) - \hat{\delta}(\tilde{\epsilon}_{xt}-\overline{\tilde{\epsilon}_x})\}^2 \\
&= T^{-1}\sum(y_t-\bar{y})^2 + \hat{\beta}'^2 T^{-1}\sum(x_{t-1}-\bar{x}_{-1})^2 + \hat{\delta}^2 T^{-1}\sum\tilde{\epsilon}_{xt}^2 \\
&\quad -2\hat{\beta}' T^{-1}\sum(y_t-\bar{y})(x_{t-1}-\bar{x}_{-1}) - 2\hat{\delta}T^{-1}\sum(y_t-\bar{y})\tilde{\epsilon}_{xt} + 2\hat{\beta}'\hat{\delta}T^{-1}\sum(x_{t-1}-\bar{x}_{-1})\tilde{\epsilon}_{xt} \\
&\quad +o_p(1) \\
&= T^{-1}\sum(y_t-\bar{y})^2 + \hat{\delta}^2 T^{-1}\sum\tilde{\epsilon}_{xt}^2 - 2\hat{\delta}T^{-1}\sum(y_t-\bar{y})\tilde{\epsilon}_{xt} + o_p(1) \\
&\Rightarrow \sigma_y^2 + (\sigma_{xy}/\sigma_x^2)^2\sigma_x^2 - 2(\sigma_{xy}/\sigma_x^2)\sigma_{xy} = \sigma_y^2(1-\rho_{xy}^2)
\end{aligned}$$

using the fact that $\overline{\tilde{\epsilon}_x} = T^{-1}\sum \tilde{\epsilon}_{xt} = T^{-1}\sum \Delta(x_t - x_1) - T\hat{\phi}T^{-2}\sum(x_{t-1} - x_1) \Rightarrow 0$. Then,

$$
\begin{aligned}
\begin{bmatrix} \hat{\beta} \\ \hat{\delta} \end{bmatrix} &= \begin{bmatrix} \sum(x_{t-1}-\bar{x}_{-1})^2 & \sum(x_{t-1}-\bar{x}_{-1})\tilde{\epsilon}_{xt} \\ \sum(x_{t-1}-\bar{x}_{-1})\tilde{\epsilon}_{xt} & \sum \tilde{\epsilon}_{xt}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum(x_{t-1}-\bar{x}_{-1})y_t \\ \sum \tilde{\epsilon}_{xt}(y_t-\bar{y}) \end{bmatrix} \\
\begin{bmatrix} T\hat{\beta} \\ \hat{\delta} \end{bmatrix} &= \left(\begin{bmatrix} T^{-1} & 0 \\ 0 & T^{-1} \end{bmatrix} \begin{bmatrix} \sum(x_{t-1}-\bar{x}_{-1})^2 & \sum(x_{t-1}-\bar{x}_{-1})\tilde{\epsilon}_{xt} \\ \sum(x_{t-1}-\bar{x}_{-1})\tilde{\epsilon}_{xt} & \sum \tilde{\epsilon}_{xt}^2 \end{bmatrix} \begin{bmatrix} T^{-1} & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \\
&\quad \cdot \begin{bmatrix} T^{-1}\sum(x_{t-1}-\bar{x}_{-1})y_t \\ T^{-1}\sum \tilde{\epsilon}_{xt}(y_t-\bar{y}) \end{bmatrix} \\
&= \begin{bmatrix} T^{-2}\sum(x_{t-1}-\bar{x}_{-1})^2 & T^{-1}\sum(x_{t-1}-\bar{x}_{-1})\tilde{\epsilon}_{xt} \\ T^{-2}\sum(x_{t-1}-\bar{x}_{-1})\tilde{\epsilon}_{xt} & T^{-1}\sum \tilde{\epsilon}_{xt}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-1}\sum(x_{t-1}-\bar{x}_{-1})y_t \\ T^{-1}\sum \tilde{\epsilon}_{xt}(y_t-\bar{y}) \end{bmatrix} \qquad \text{(A.1)} \\
&\Rightarrow \begin{bmatrix} \sigma_x^2\int \bar{W}_{1c}(r)^2dr & \sigma_x^2\int \bar{W}_{1c}(r)dW_{1c}(r) - \sigma_x^2 L(c)\int \bar{W}_{1c}(r)^2dr \\ 0 & \sigma_x^2 \end{bmatrix}^{-1} \\
&\quad \cdot \begin{bmatrix} g\sigma_x^2\int \bar{W}_{1c}(r)^2dr + \sigma_x\sigma_y\int \bar{W}_{1c}(r)d\left\{\rho_{xy}W_1(r) + \sqrt{1-\rho_{xy}^2}W_2(r)\right\} \\ \sigma_{xy} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sigma_x^2\int \bar{W}_{1c}(r)^2dr} & -\frac{\int \bar{W}_{1c}(r)dW_{1c}(r) - L(c)\int \bar{W}_{1c}(r)^2dr}{\sigma_x^2\int \bar{W}_{1c}(r)^2dr} \\ 0 & 1/\sigma_x^2 \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} g\sigma_x^2\int \bar{W}_{1c}(r)^2dr + \sigma_x\sigma_y\int \bar{W}_{1c}(r)d\left\{\rho_{xy}W_1(r) + \sqrt{1-\rho_{xy}^2}W_2(r)\right\} \\ \sigma_{xy} \end{bmatrix} \\
&= \begin{bmatrix} \frac{g\sigma_x^2\int \bar{W}_{1c}(r)^2dr + \sigma_x\sigma_y\int \bar{W}_{1c}(r)d\left\{\rho_{xy}W_1(r) + \sqrt{1-\rho_{xy}^2}W_2(r)\right\}}{\sigma_x^2\int \bar{W}_{1c}(r)^2dr} - \sigma_{xy}\frac{\int \bar{W}_{1c}(r)dW_{1c}(r) - L(c)\int \bar{W}_{1c}(r)^2dr}{\sigma_x^2\int \bar{W}_{1c}(r)^2dr} \\ \sigma_{xy}/\sigma_x^2 \end{bmatrix}
\end{aligned}
$$

The expression for $T\hat{\beta}$ can be simplified as follows:

$$
\begin{aligned}
T\hat{\beta} &\Rightarrow g + \frac{\sigma_x\sigma_y\int \bar{W}_{1c}(r)d\left\{\rho_{xy}W_1(r) + \sqrt{1-\rho_{xy}^2}W_2(r)\right\}}{\sigma_x^2\int \bar{W}_{1c}(r)^2dr} - \sigma_{xy}\frac{\int \bar{W}_{1c}(r)dW_{1c}(r) - L(c)\int \bar{W}_{1c}(r)^2dr}{\sigma_x^2\int \bar{W}_{1c}(r)^2dr} \\
&= g + \frac{\sigma_x\sigma_y\int \bar{W}_{1c}(r)d\left\{\rho_{xy}W_1(r) + \sqrt{1-\rho_{xy}^2}W_2(r)\right\}}{\sigma_x^2\int \bar{W}_{1c}(r)^2dr} - \sigma_{xy}\left(\frac{\int \bar{W}_{1c}(r)dW_{1c}(r)}{\sigma_x^2\int \bar{W}_{1c}(r)^2dr} - \frac{L(c)}{\sigma_x^2}\right) \\
&= g + \frac{\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}\int \bar{W}_{1c}(r)dW_2(r)}{\sigma_x^2\int \bar{W}_{1c}(r)^2dr} + \frac{\sigma_x\sigma_y\rho_{xy}\int \bar{W}_{1c}(r)dW_1(r)}{\sigma_x^2\int \bar{W}_{1c}(r)^2dr} \\
&\quad -\frac{\sigma_{xy}}{\sigma_x^2}\left(c + \frac{\int \bar{W}_{1c}(r)dW_1(r)}{\int \bar{W}_{1c}(r)^2dr} - L(c)\right) \\
&= g + \frac{\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}\int \bar{W}_{1c}(r)dW_2(r)}{\sigma_x^2\int \bar{W}_{1c}(r)^2dr} - \frac{\sigma_{xy}}{\sigma_x^2}\left(c - c - \frac{\int W_{1c}(r)dW_1(r)}{\int W_{1c}(r)^2dr}\right) \\
&= g + \frac{\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}\int \bar{W}_{1c}(r)dW_2(r)}{\sigma_x^2\int \bar{W}_{1c}(r)^2dr} + \frac{\sigma_{xy}\int W_{1c}(r)dW_1(r)}{\sigma_x^2\int W_{1c}(r)^2dr}
\end{aligned}
$$

Hence, using m to denote element [1,1] of the inverse matrix in (A.1), we find

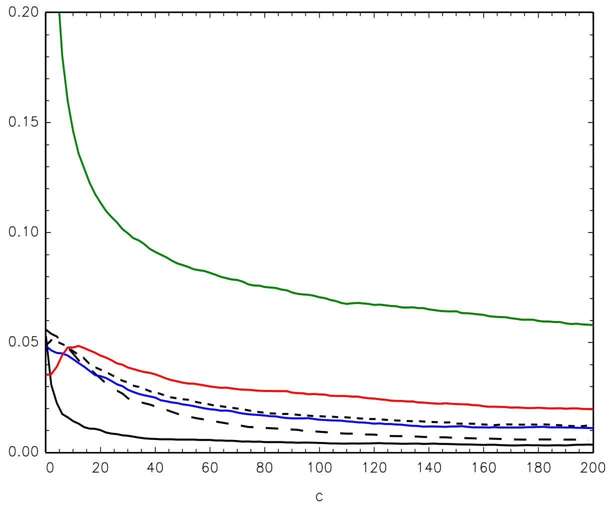
$$
\begin{aligned}
t^* &= \frac{T\hat{\beta}}{\sqrt{T^{-1}\sum \hat{v}_t^2 T^2 m}} \\
&\Rightarrow \frac{g + \frac{\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}\int \bar{W}_{1c}(r)dW_2(r)}{\sigma_x^2\int \bar{W}_{1c}(r)^2 dr} + \frac{\sigma_{xy}\int W_{1c}(r)dW_1(r)}{\sigma_x^2\int W_{1c}(r)^2 dr}}{\sqrt{\frac{\sigma_y^2(1-\rho_{xy}^2)}{\sigma_x^2\int \bar{W}_{1c}(r)^2 dr}}} \\
&= \left[g + \frac{\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}\int \bar{W}_{1c}(r)dW_2(r)}{\sigma_x^2\int \bar{W}_{1c}(r)^2 dr} + \frac{\sigma_{xy}\int W_{1c}(r)dW_1(r)}{\sigma_x^2\int W_{1c}(r)^2 dr}\right]\sqrt{\frac{\sigma_x^2\int \bar{W}_{1c}(r)^2 dr}{\sigma_y^2(1-\rho_{xy}^2)}} \\
&= \frac{g\sigma_x}{\sigma_y}\frac{\sqrt{\int \bar{W}_{1c}(r)^2 dr}}{\sqrt{1-\rho_{xy}^2}} + \frac{\int \bar{W}_{1c}(r)dW_2(r)}{\sqrt{\int \bar{W}_{1c}(r)^2 dr}} + \frac{\rho_{xy}\sqrt{\int \bar{W}_{1c}(r)^2 dr}\int W_{1c}(r)dW_1(r)}{\sqrt{1-\rho_{xy}^2}\int W_{1c}(r)^2 dr}
\end{aligned}
$$

Proofs of Theorem 2, 3 and 4: The proofs of Theorems 2, 3 and 4 are entirely straightforward and are therefore omitted, but are available from the authors on request.

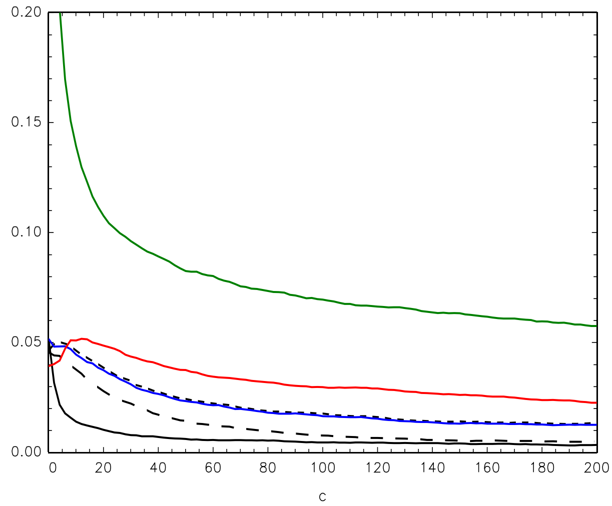
Table 2. Application to monthly U.S. stock index returns, 1970:1-2017:12.

Predictor	$\hat{\rho}_{xy}$	p^{DF}	t_C	t'_C	t^*_C	t^w_C	t^{wt}_C	BD	IVX_1	IVX_2	Q
d/p	−0.99	0.48	1.34	1.06	4.02	1.74	$1.74^{\dagger}$	0.06	1.12	1.26	
d/y	−0.04	0.40	$1.44^{\dagger}$	$1.25^{\dagger}$	$1.42^{\dagger}$	$1.27^{\dagger}$	1.27	0.00	1.26	1.59	
dfr	0.24	0.00	$1.62^{\dagger\dagger}$	1.10	$4.36^{\dagger\dagger}$	1.10	$1.62^{\dagger\dagger}$	1.51	$1.64^{\dagger}$	2.68	††
inf	−0.07	0.01	−1.01	−0.99	−0.95	−0.99	−0.99	1.26	−1.11	1.24	††
$svar$	−0.31	0.00	$-2.99^{\dagger\dagger}$	$-2.88^{\dagger\dagger}$	$-3.42^{\dagger\dagger}$	$-2.88^{\dagger\dagger}$	$-2.99^{\dagger\dagger}$	$12.20^{\dagger\dagger}$	$-3.03^{\dagger\dagger}$	$9.19^{\dagger\dagger}$	††

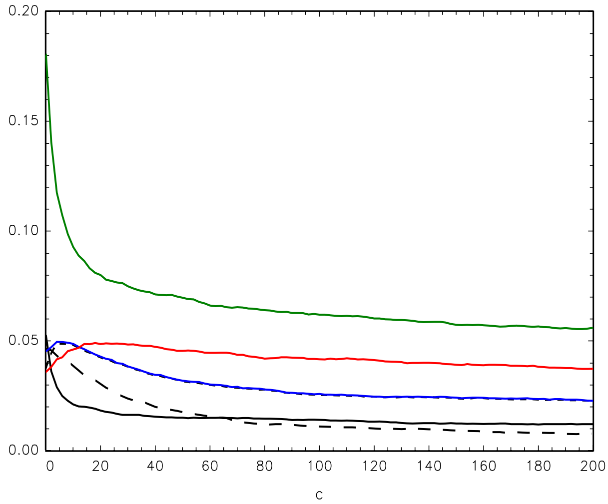
Note: † and †† denote rejection at the 0.10-level and 0.05-level respectively.



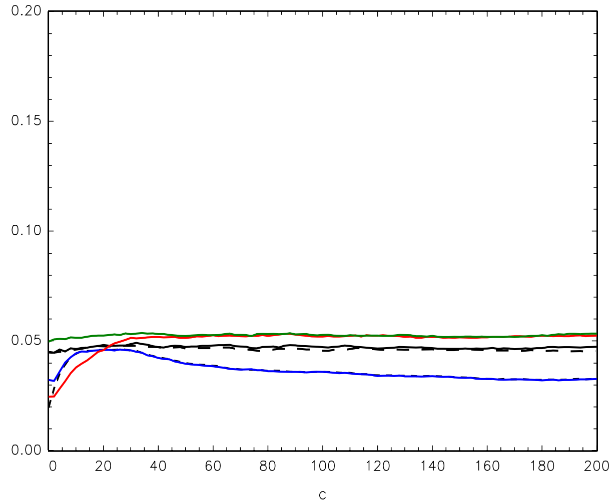
(a) $\rho_{xy} = -0.95$



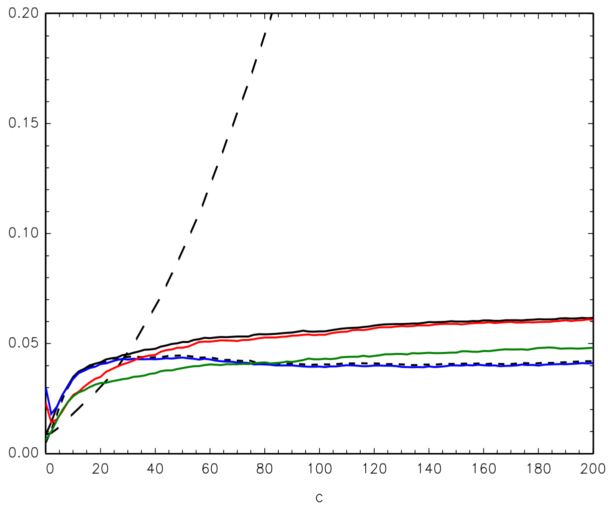
(b) $\rho_{xy} = -0.9$



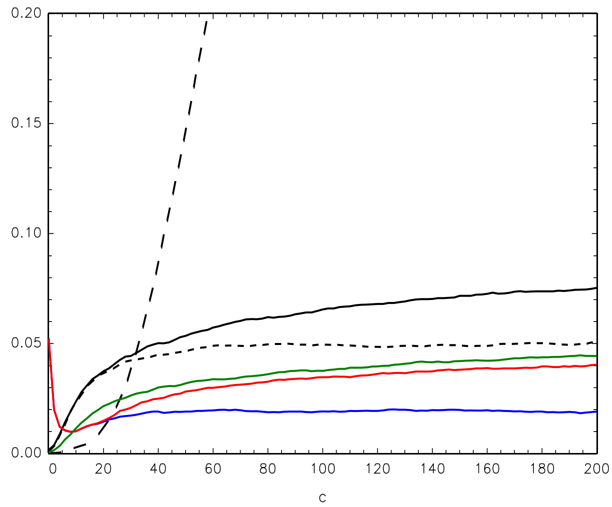
(c) $\rho_{xy} = -0.5$



(d) $\rho_{xy} = 0$

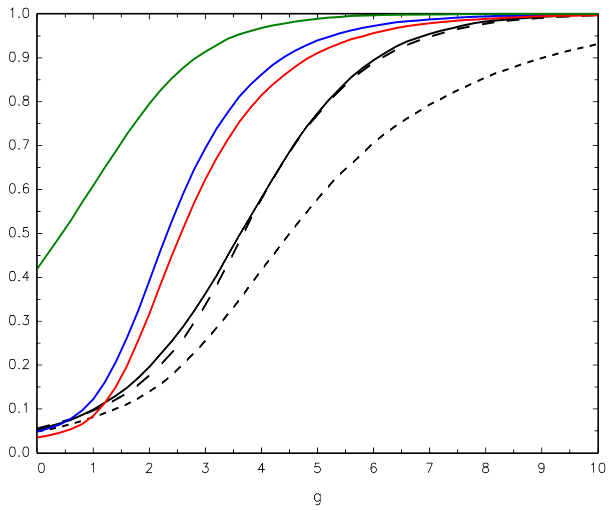


(e) $\rho_{xy} = 0.5$

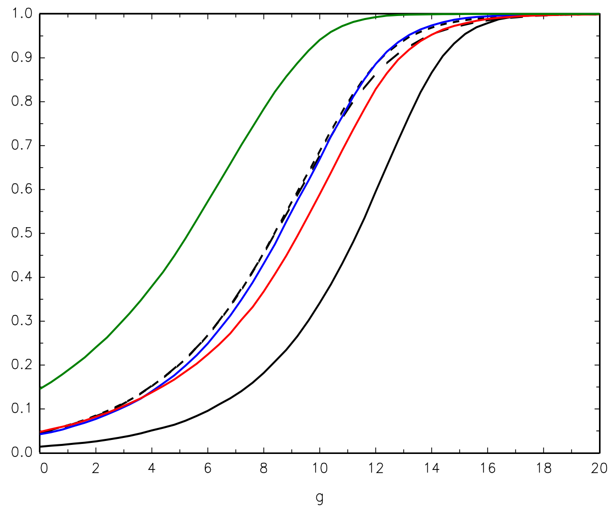


(f) $\rho_{xy} = 0.9$

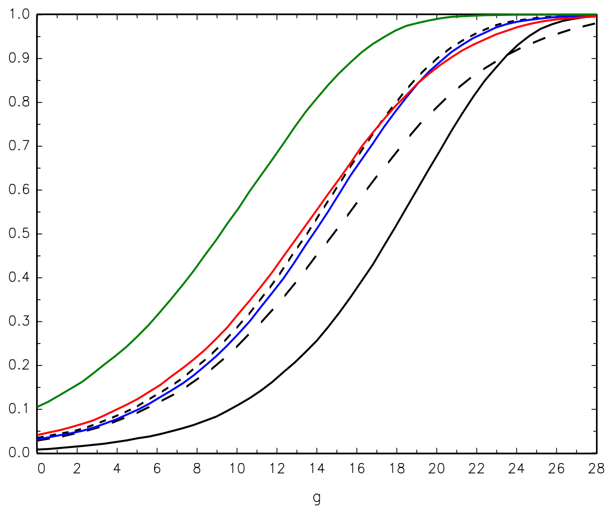
Figure 1. Finite sample size of nominal 0.05-level tests, $T = 200$;
t_N: green solid line, t_C: black solid line, t'_C: short-dashed line, t^*_C: long-dashed line, t^w_C: blue solid line, t^{wt}_C: red solid line



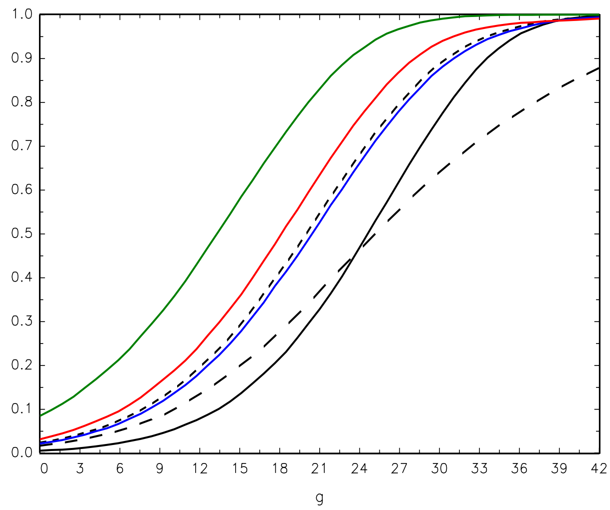
(a) $c = 0$



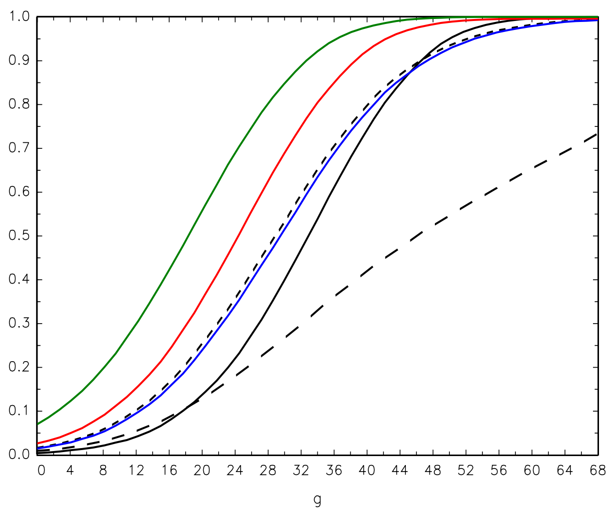
(b) $c = 10$



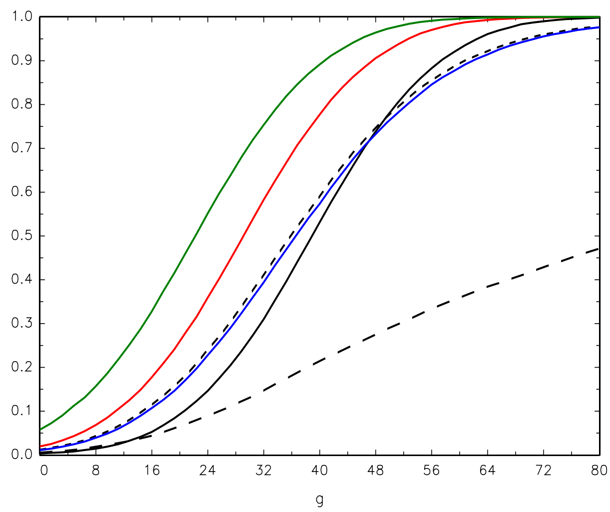
(c) $c = 25$



(d) $c = 50$

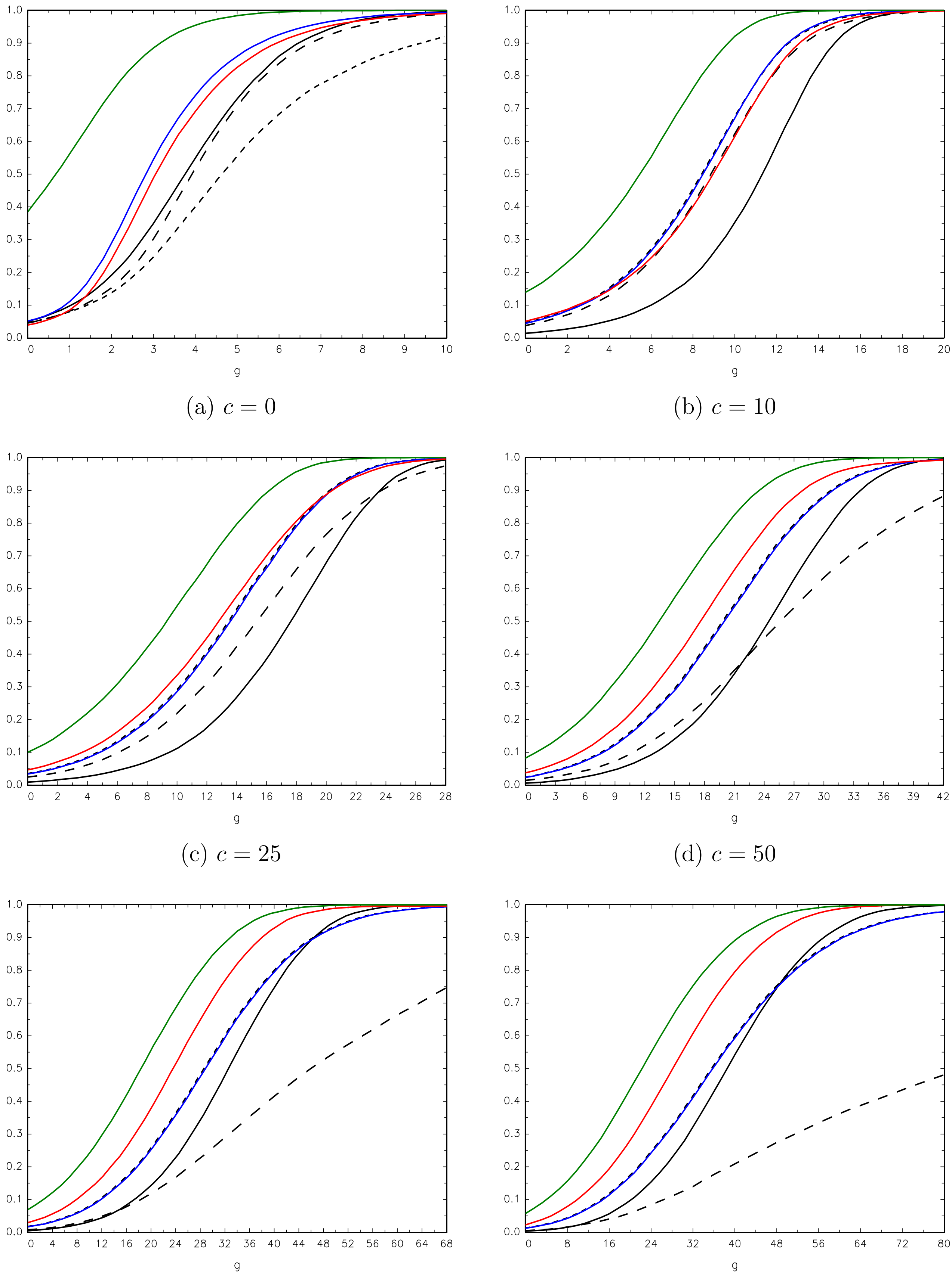


(e) $c = 100$



(f) $c = 200$

Figure 2. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.95$;
t_N: green line, t_C: solid black line, t'_C: short-dashed line, t^*_C: long-dashed line, t^w_C: blue line, t^{wt}_C: red line

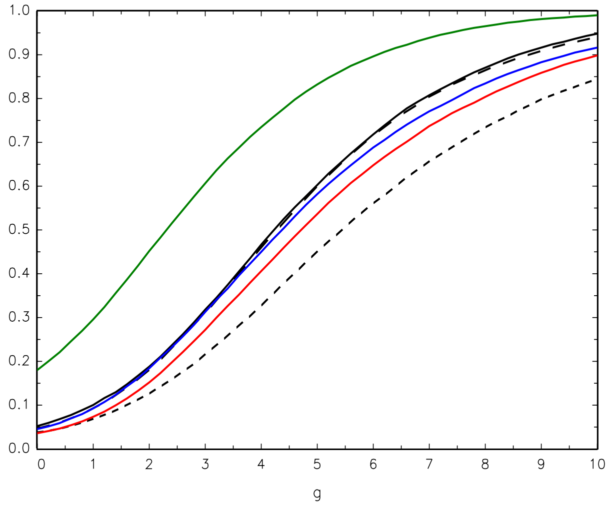


(a) $c = 0$ (b) $c = 10$

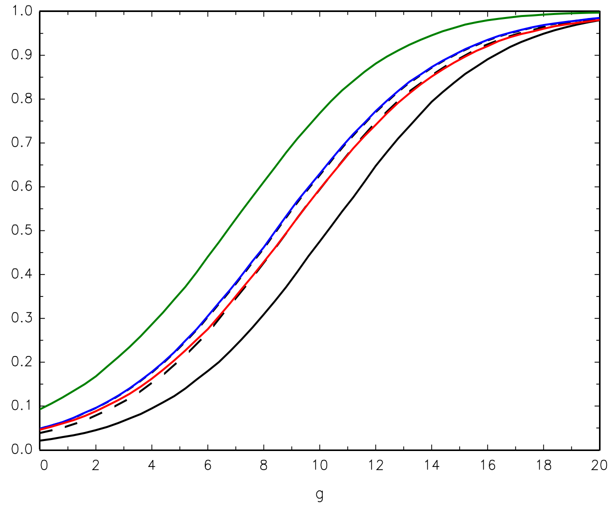
(c) $c = 25$ (d) $c = 50$

(e) $c = 100$ (f) $c = 200$

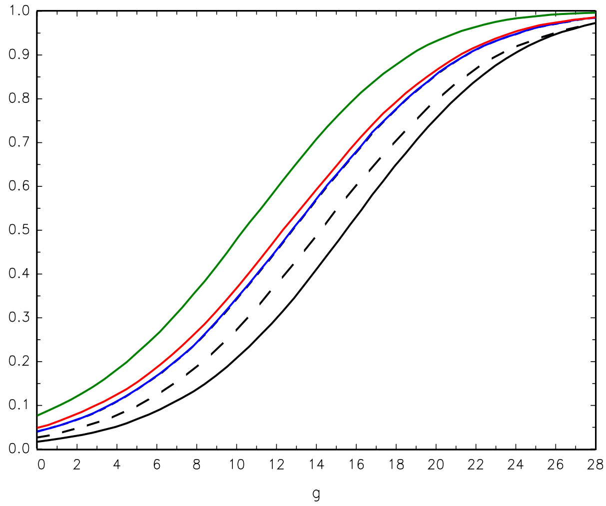
Figure 3. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.9$;
t_N: green solid line, t_C: black solid line, t'_C: short-dashed line, t^*_C: long-dashed line, t^w_C: blue solid line, t^{wt}_C: red solid line



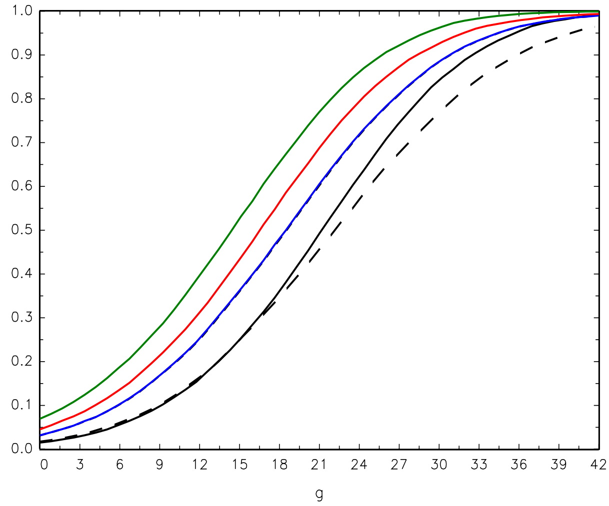
(a) $c = 0$



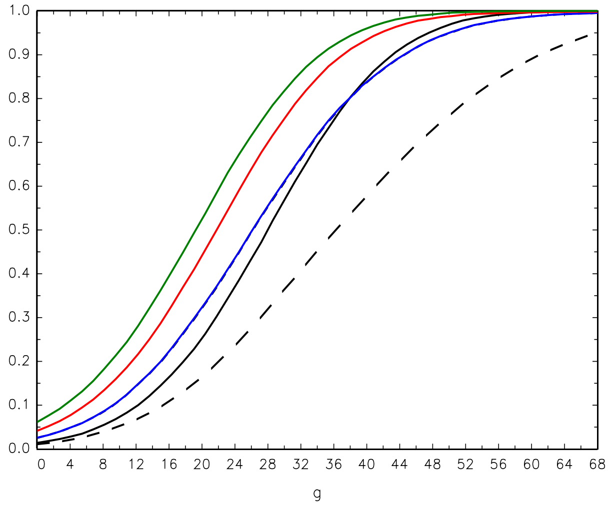
(b) $c = 10$



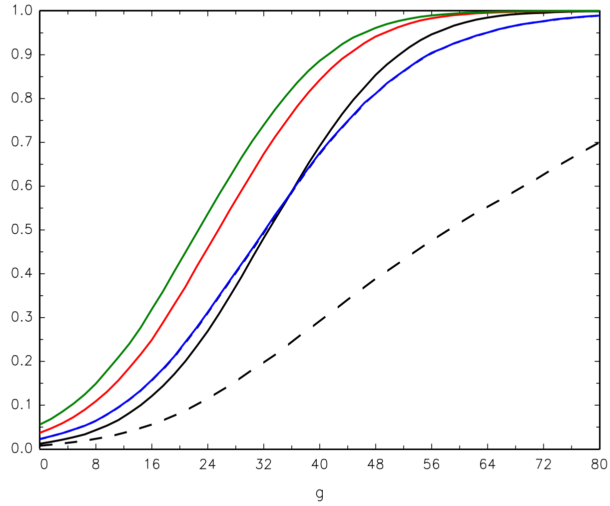
(c) $c = 25$



(d) $c = 50$

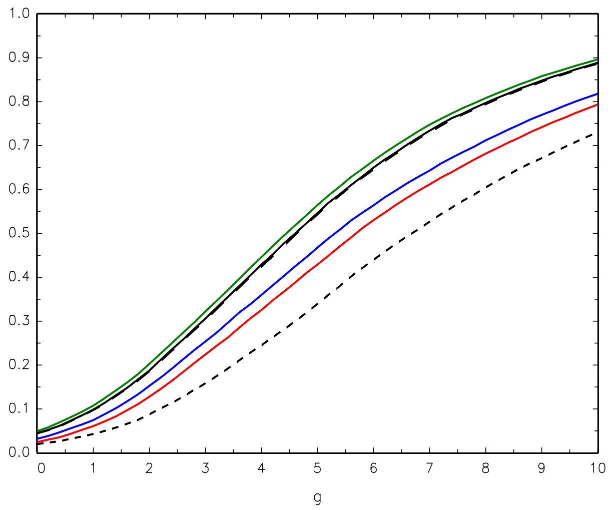


(e) $c = 100$

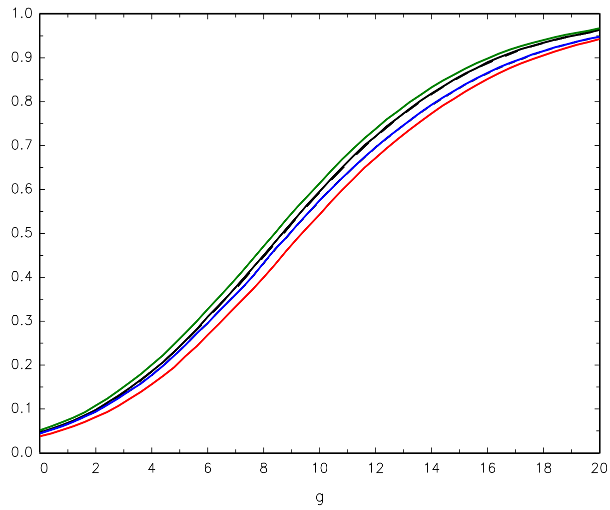


(f) $c = 200$

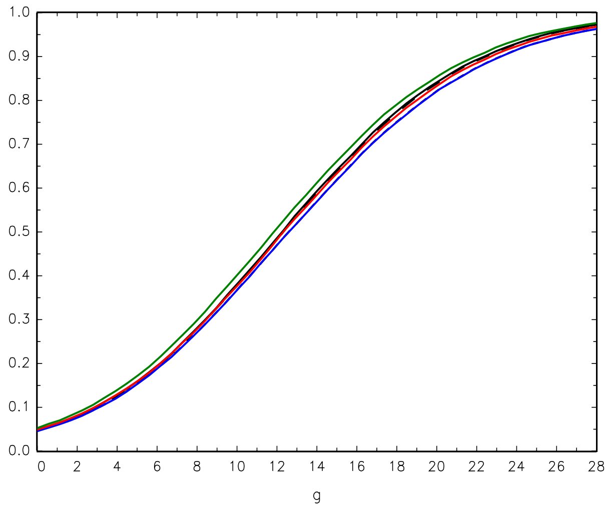
Figure 4. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.5$;
t_N: green solid line, t_C: black solid line, t'_C: short-dashed line, t^*_C: long-dashed line, t^w_C: blue solid line, t^{wt}_C: red solid line



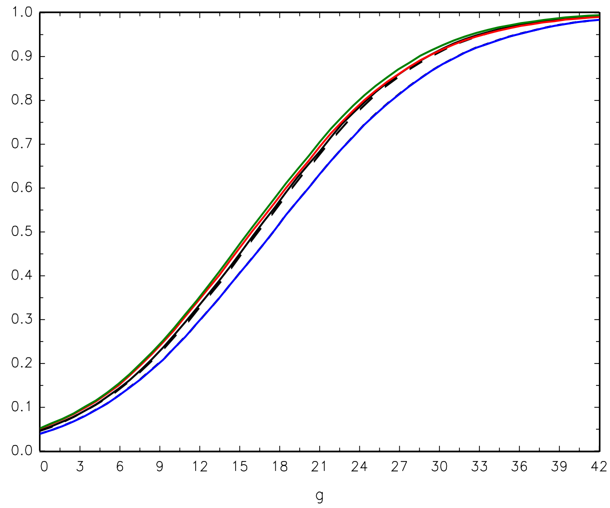
(a) $c = 0$



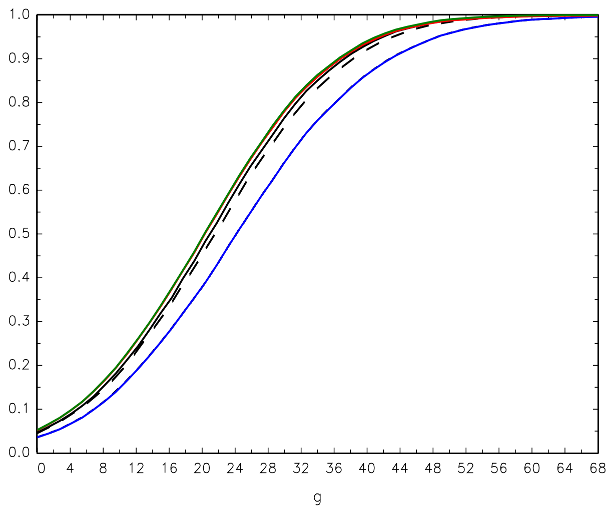
(b) $c = 10$



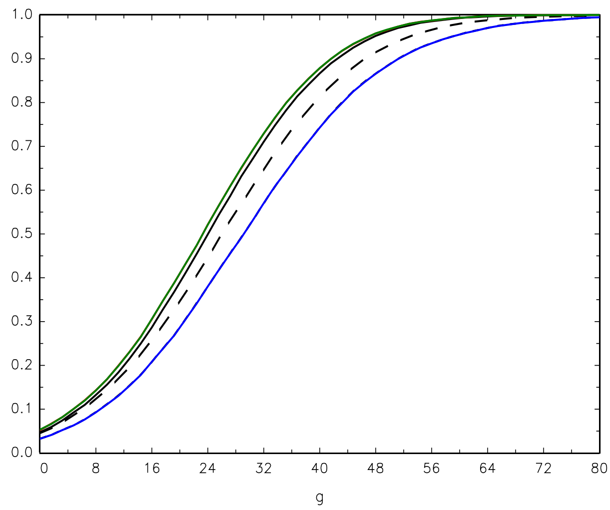
(c) $c = 25$



(d) $c = 50$

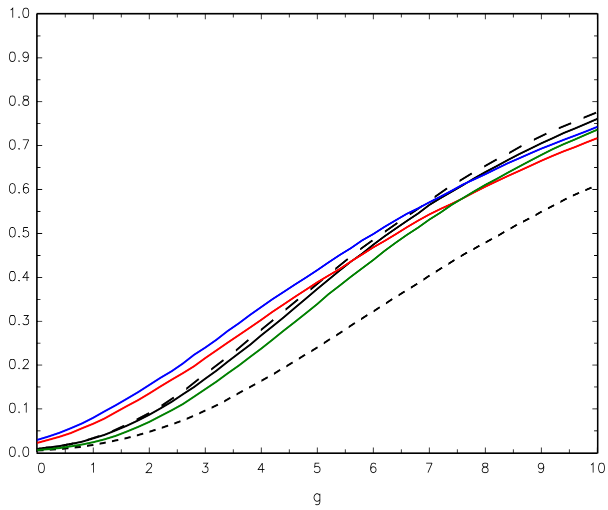


(e) $c = 100$

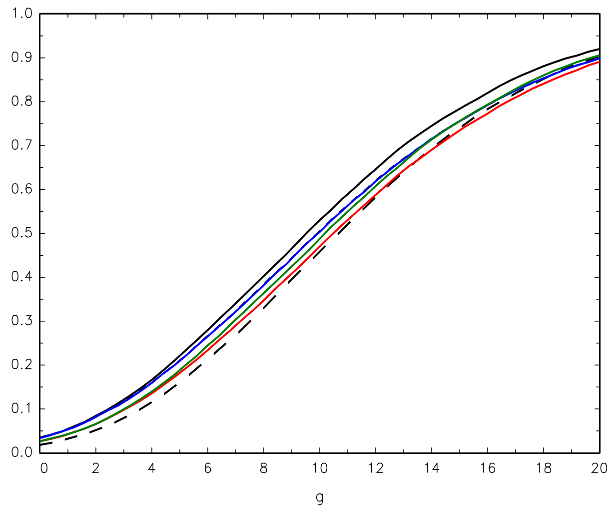


(f) $c = 200$

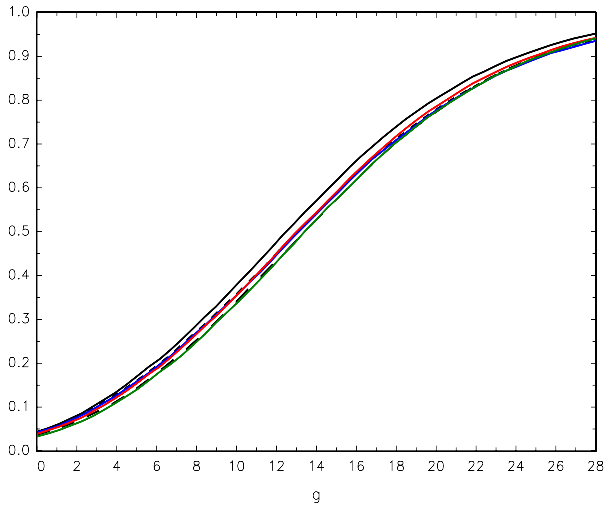
Figure 5. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0$; t_N: —— , t_C: —— , t'_C: - - - , t^*_C: — — , t^w_C: —— , t^{wt}_C: ——



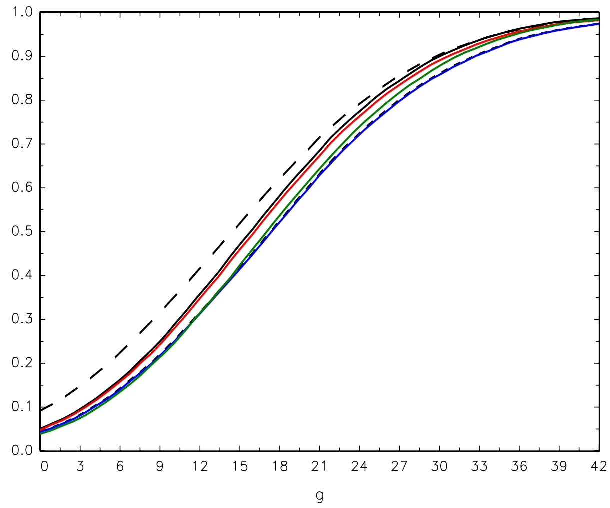
(a) $c = 0$



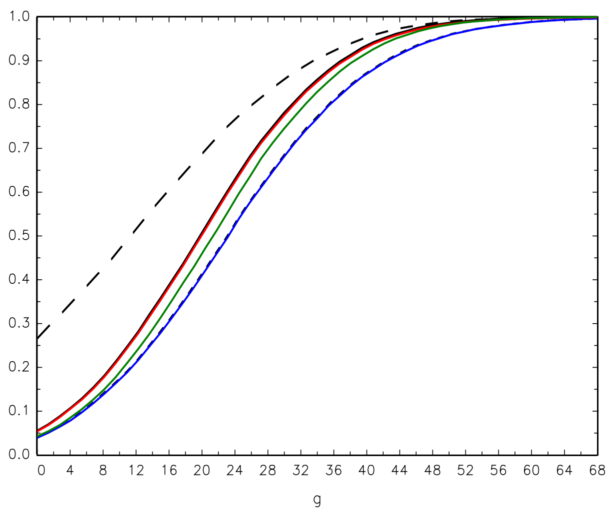
(b) $c = 10$



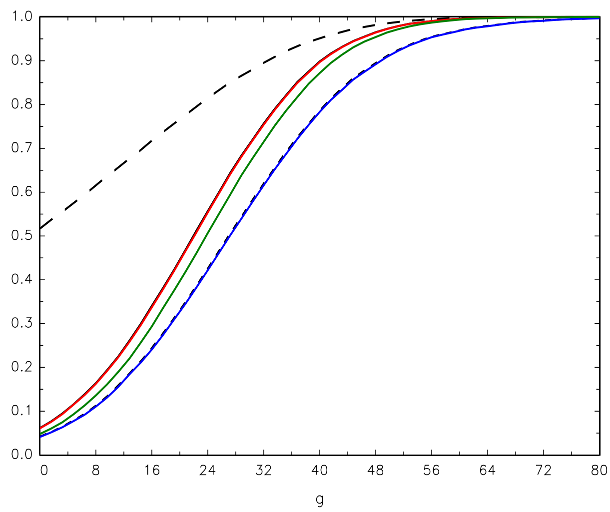
(c) $c = 25$



(d) $c = 50$

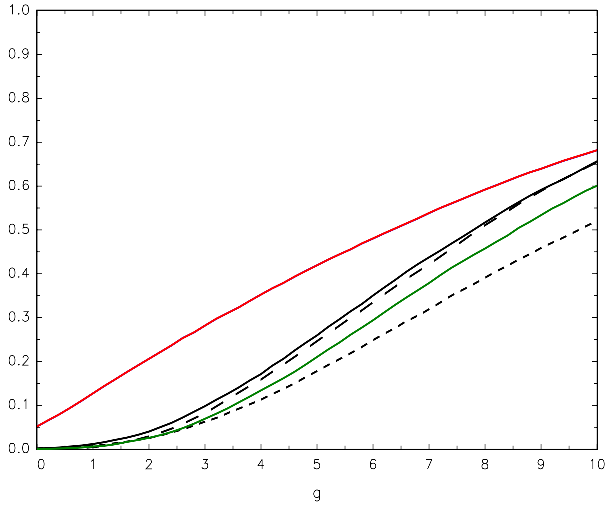


(e) $c = 100$

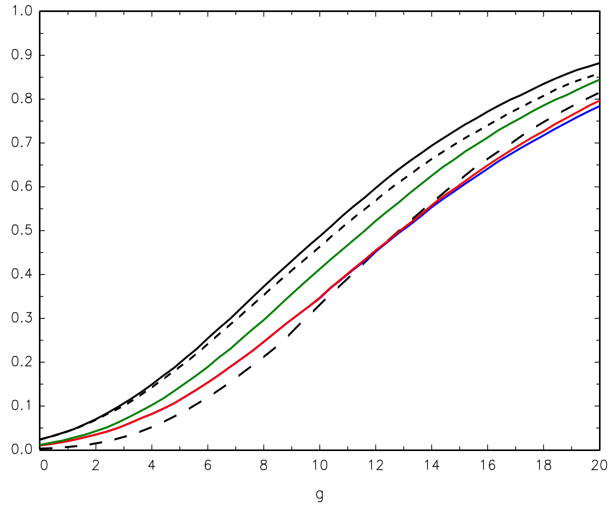


(f) $c = 200$

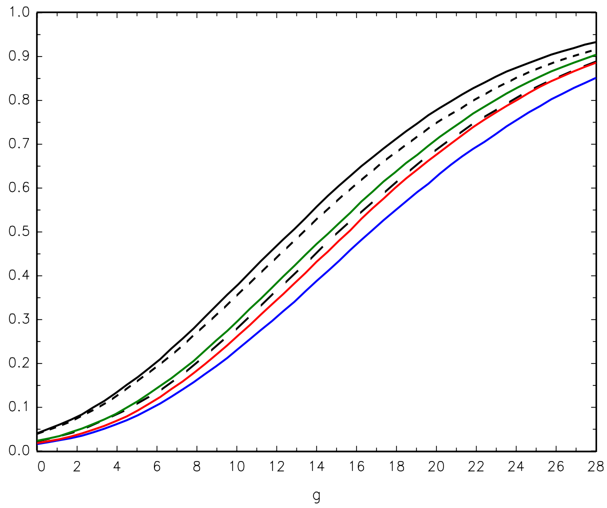
Figure 6. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.5$;
t_N: —— , t_C: —— , t'_C: - - - , t^*_C: — — , t^w_C: —— , t^{wt}_C: ——



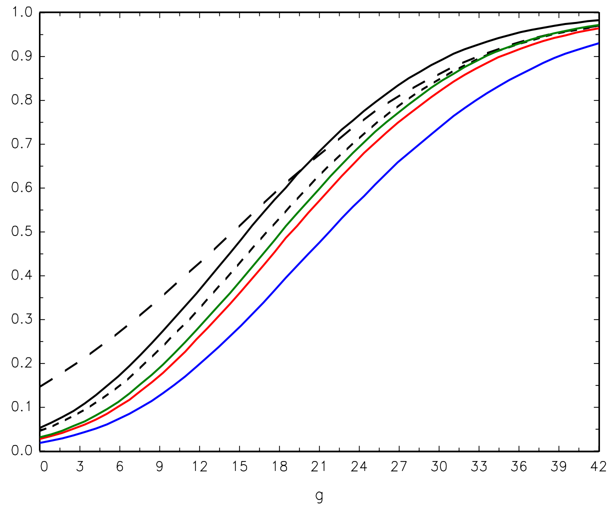
(a) $c = 0$



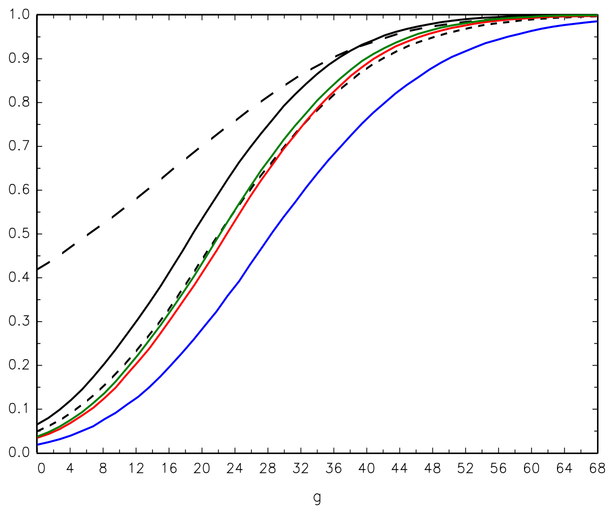
(b) $c = 10$



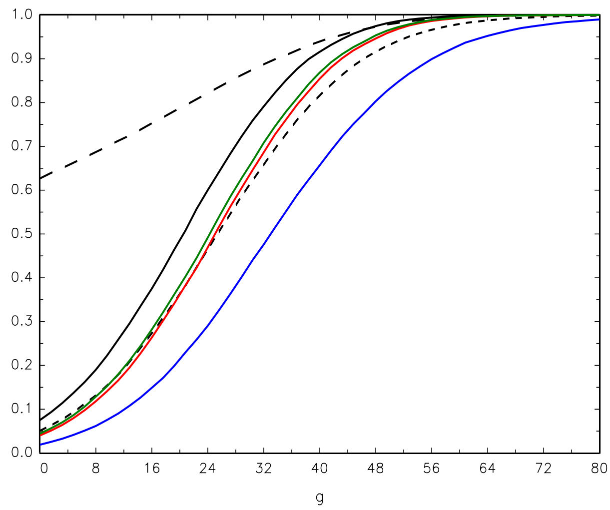
(c) $c = 25$



(d) $c = 50$

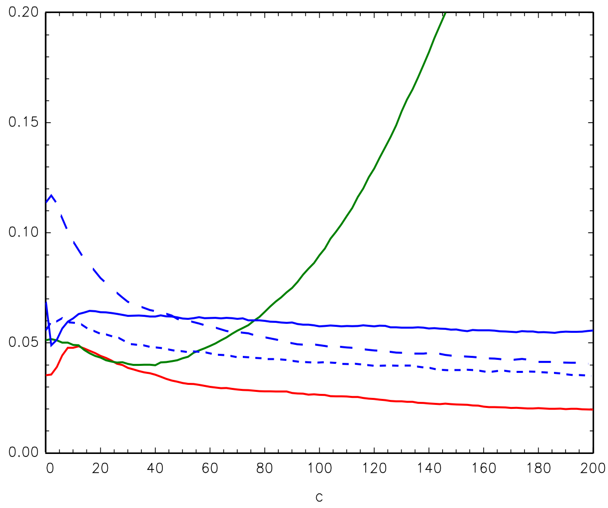


(e) $c = 100$

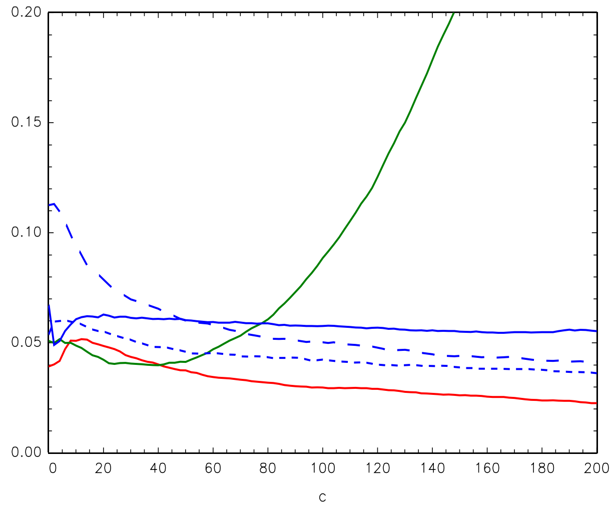


(f) $c = 200$

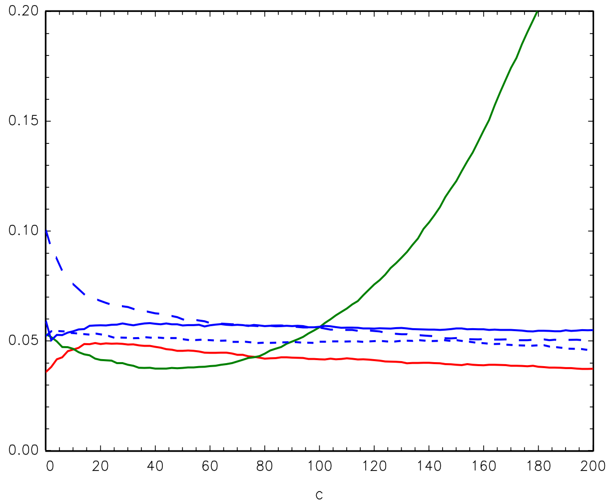
Figure 7. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.9$;
t_N: green solid, t_C: black solid, t'_C: short dashed, t^*_C: long dashed, t^w_C: blue solid, t^{wt}_C: red solid



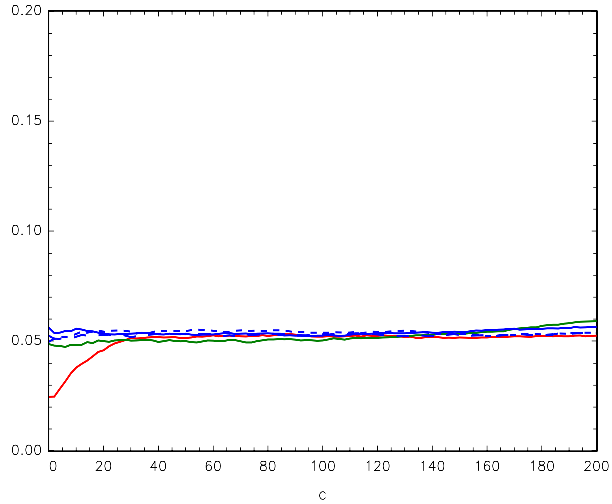
(a) $\rho_{xy} = -0.95$



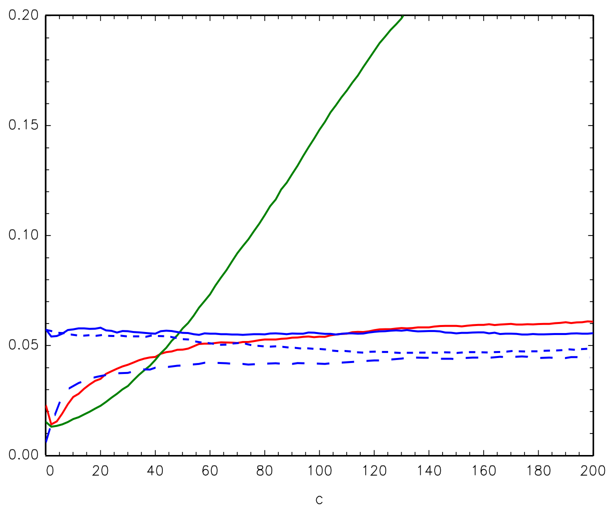
(b) $\rho_{xy} = -0.9$



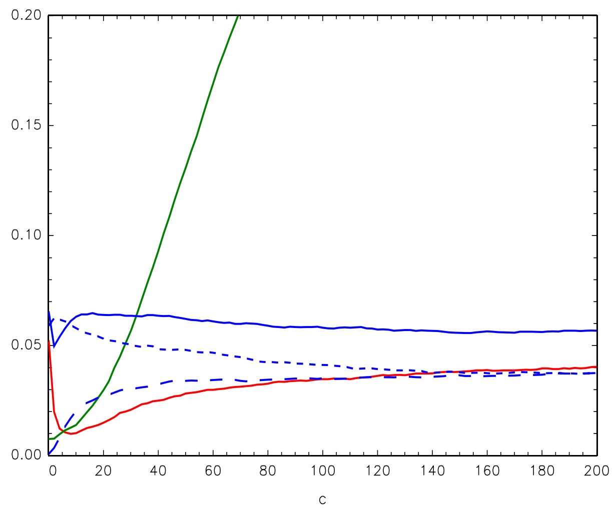
(c) $\rho_{xy} = -0.5$



(d) $\rho_{xy} = 0$



(e) $\rho_{xy} = 0.5$



(f) $\rho_{xy} = 0.9$

Figure 8. Finite sample size of nominal 0.05-level tests, $T = 200$;
t_C^{wt}: , Q: , BD: , IVX_1: , IVX_2:

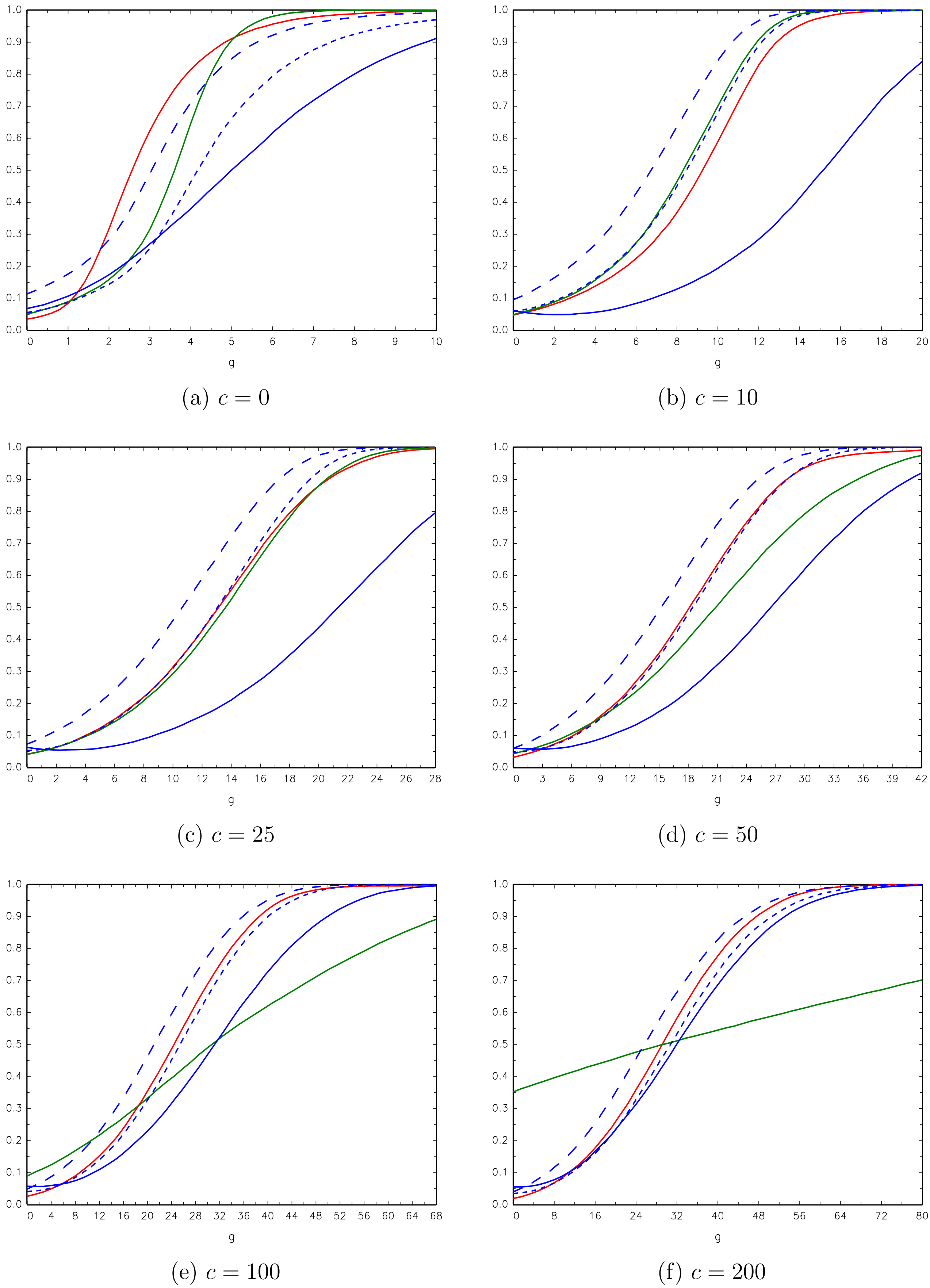
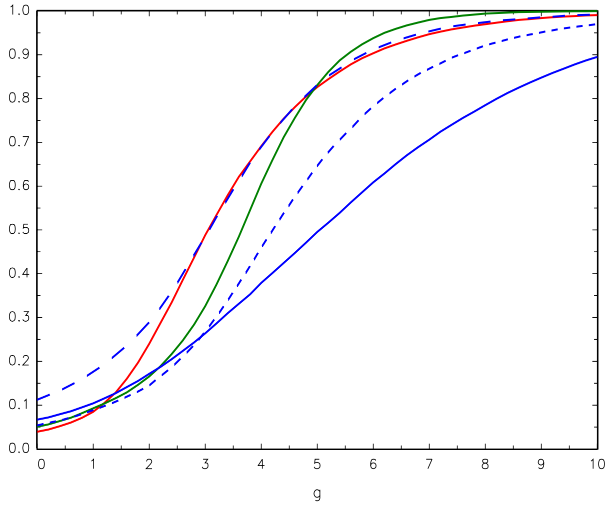
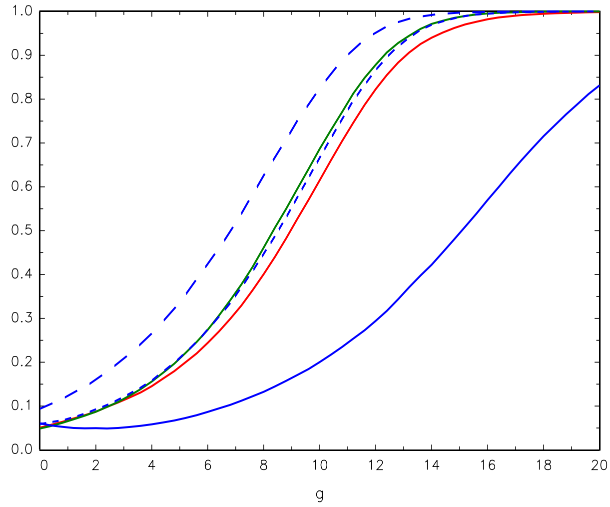


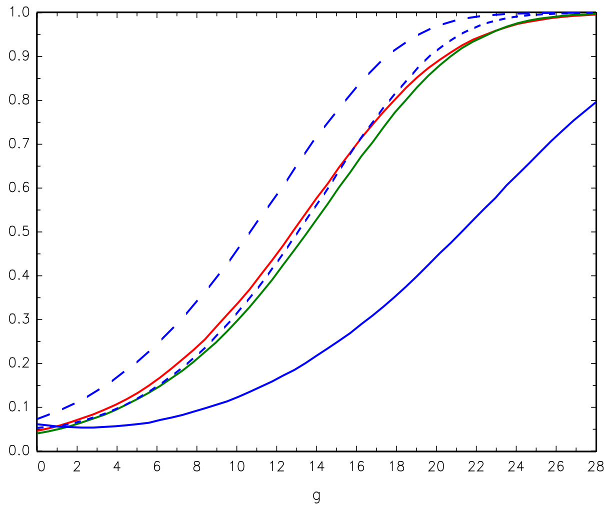
Figure 9. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.95$;
t_C^{wt}: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line



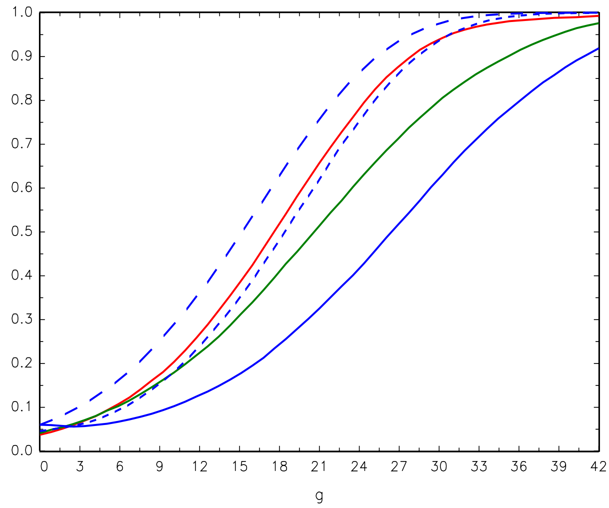
(a) $c = 0$



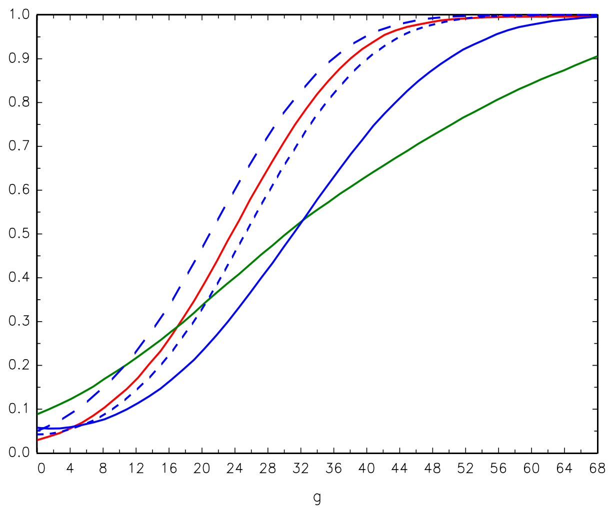
(b) $c = 10$



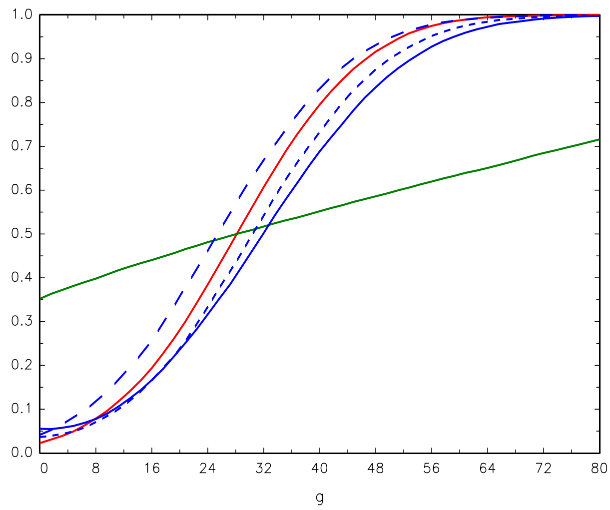
(c) $c = 25$



(d) $c = 50$

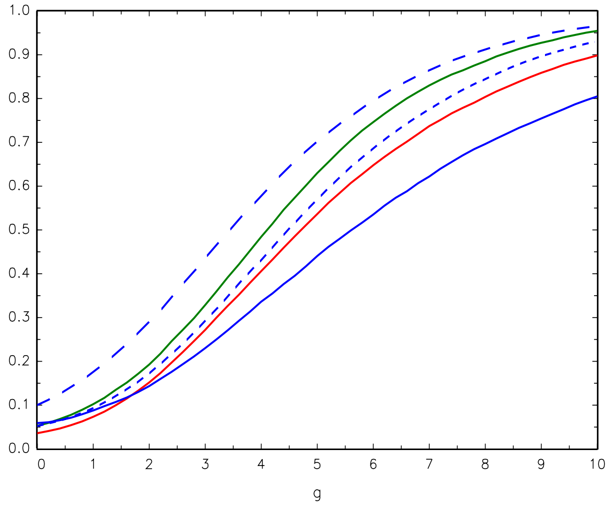


(e) $c = 100$

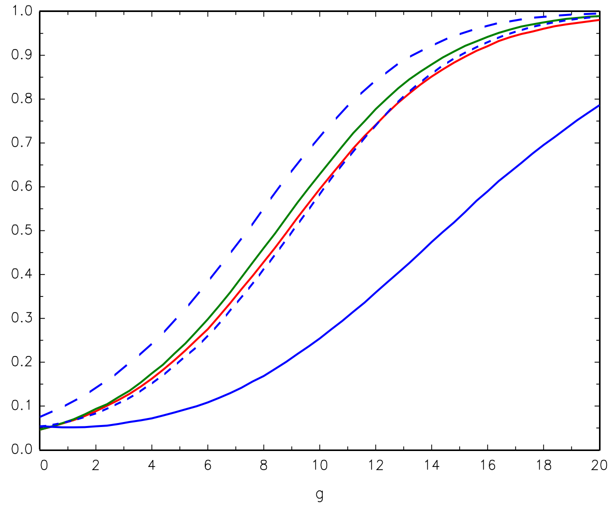


(f) $c = 200$

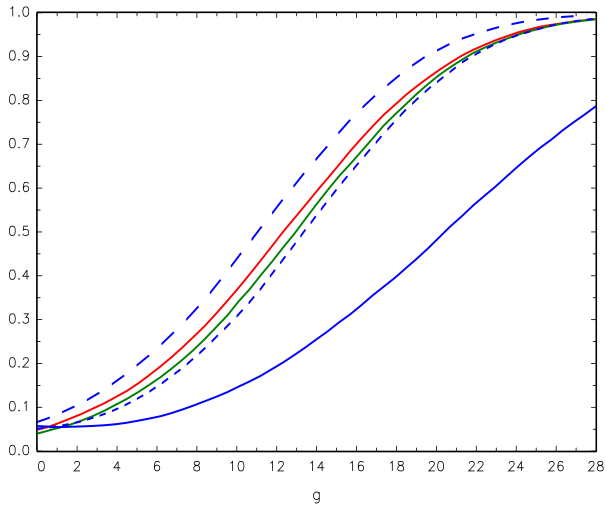
Figure 10. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.9$;
t_C^{wt}: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line



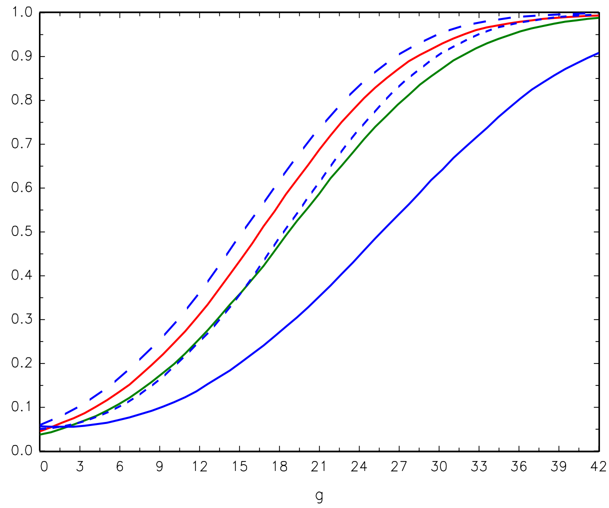
(a) $c = 0$



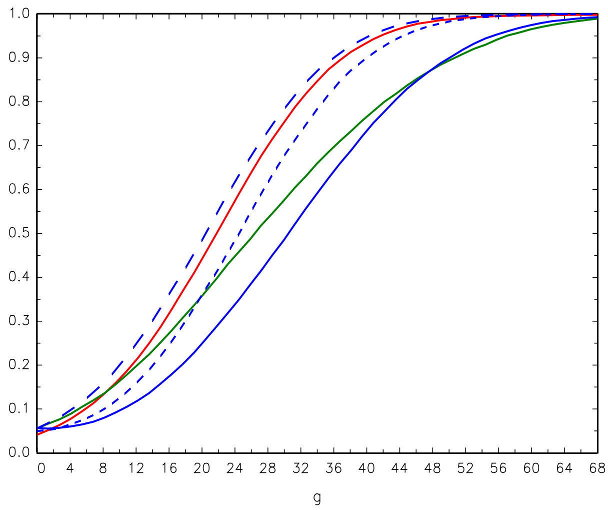
(b) $c = 10$



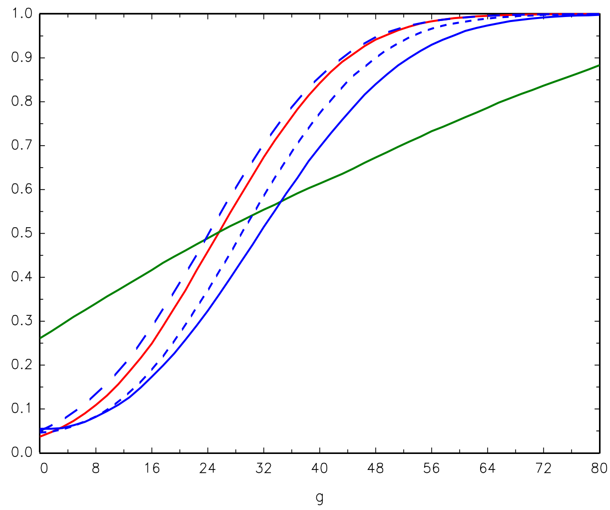
(c) $c = 25$



(d) $c = 50$

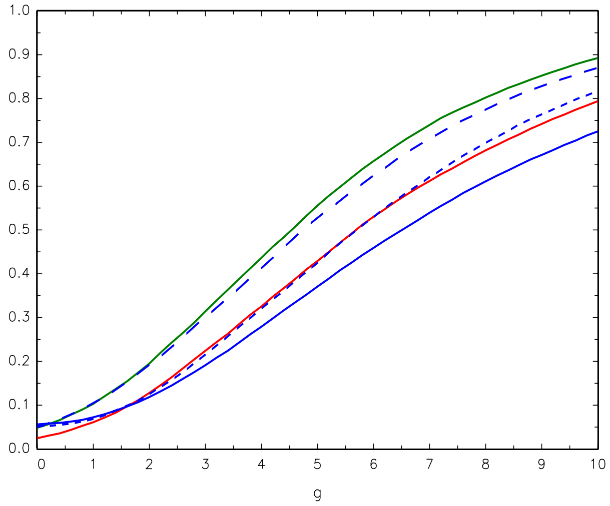


(e) $c = 100$

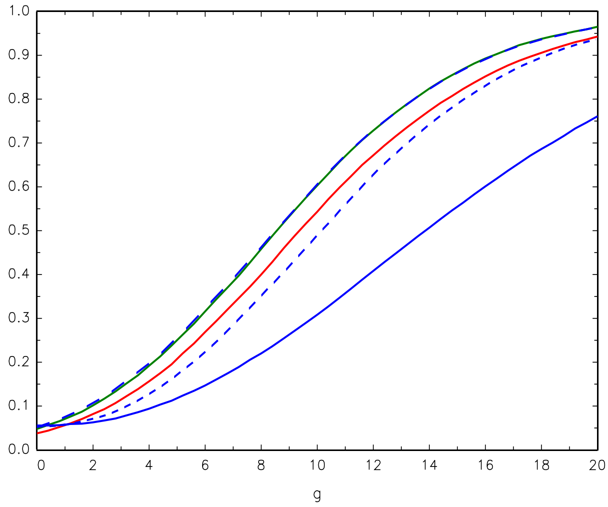


(f) $c = 200$

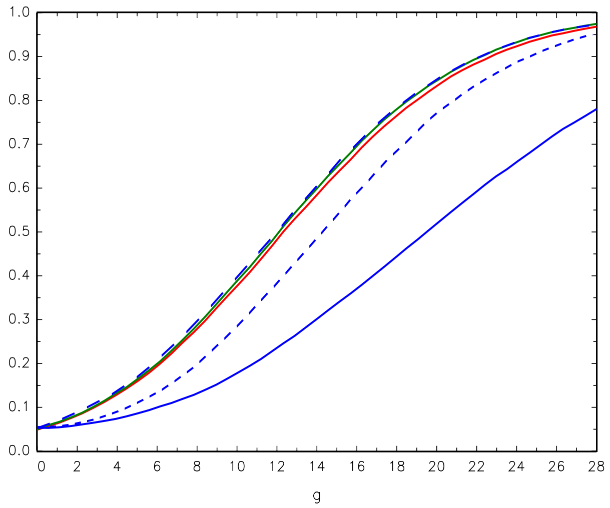
Figure 11. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.5$; t_C^{wt}: —— , Q: —— , BD: —— , IVX_1: — — , IVX_2: - - -



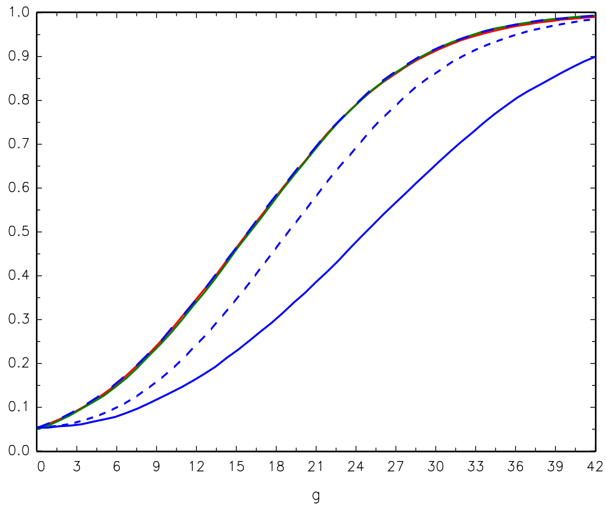
(a) $c = 0$



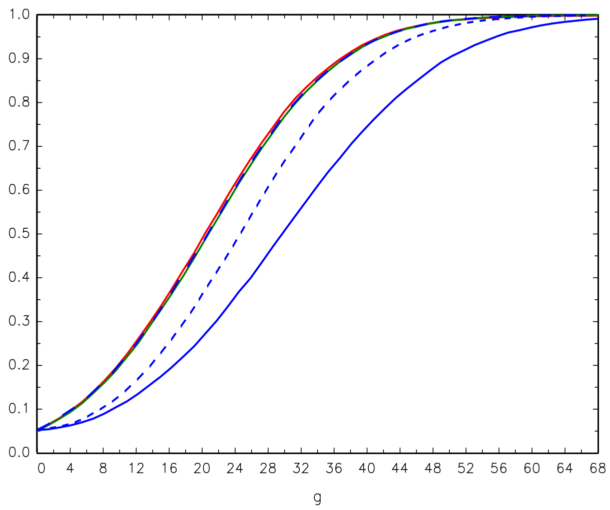
(b) $c = 10$



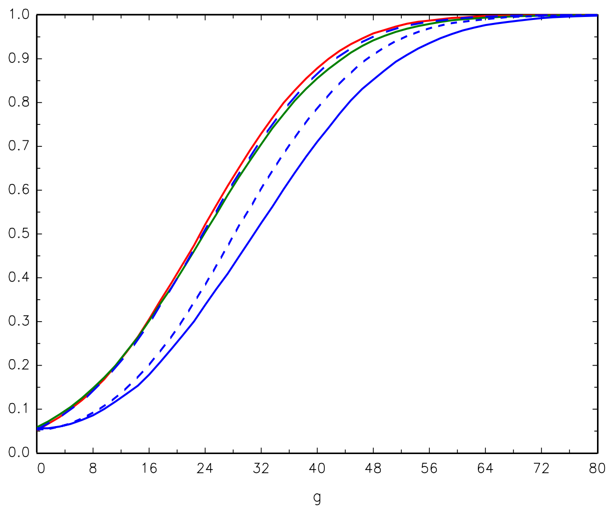
(c) $c = 25$



(d) $c = 50$

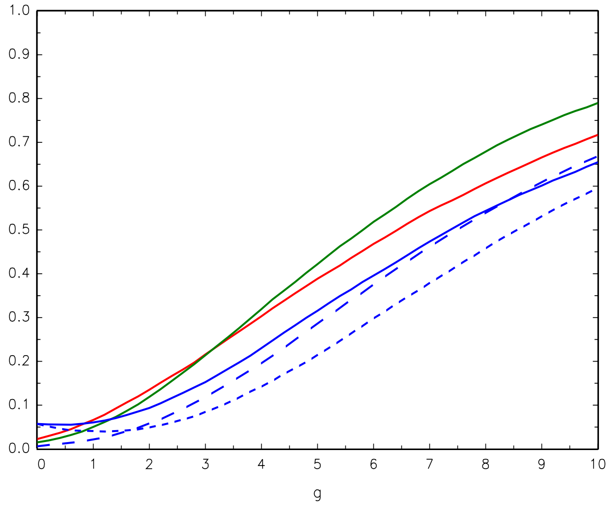


(e) $c = 100$

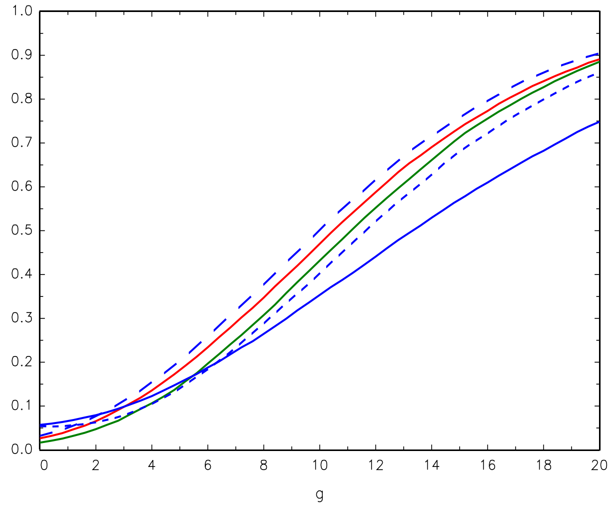


(f) $c = 200$

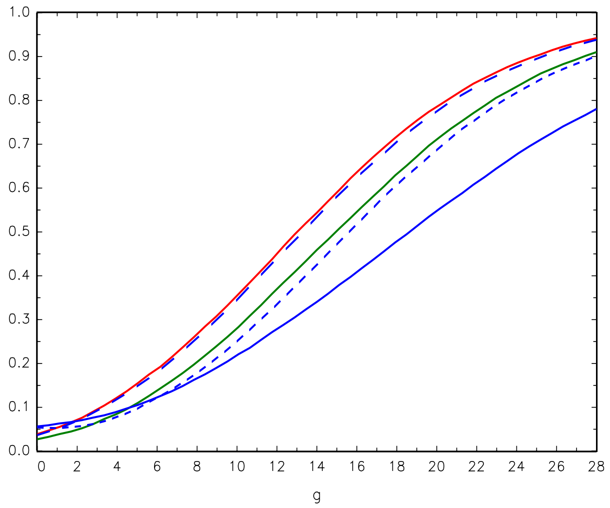
Figure 12. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0$; t_C^{wt}: ——— , Q: ——— , BD: ——— , IVX_1: — — — , IVX_2: - - -



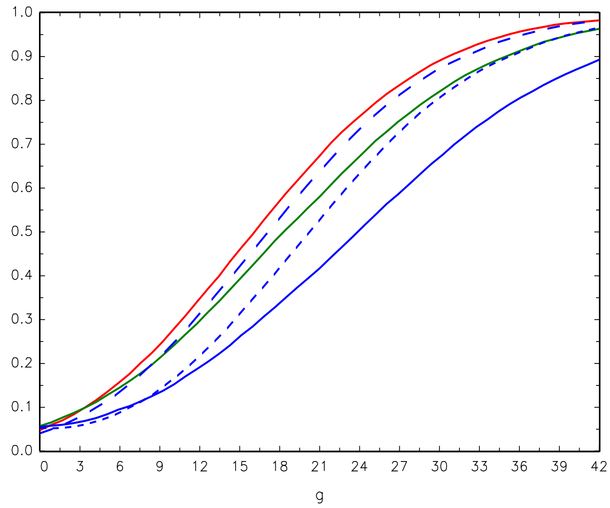
(a) $c = 0$



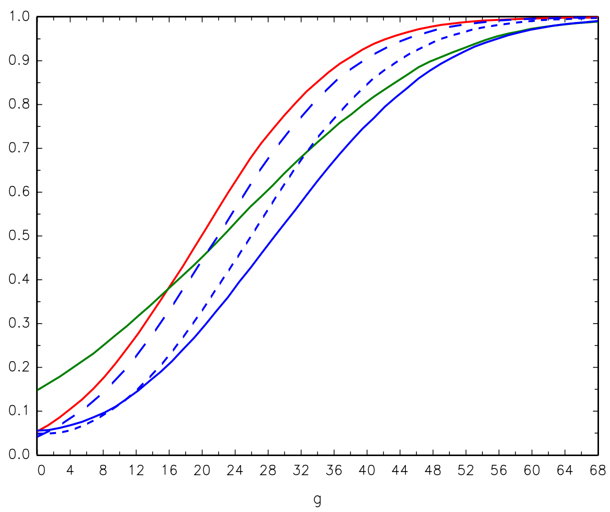
(b) $c = 10$



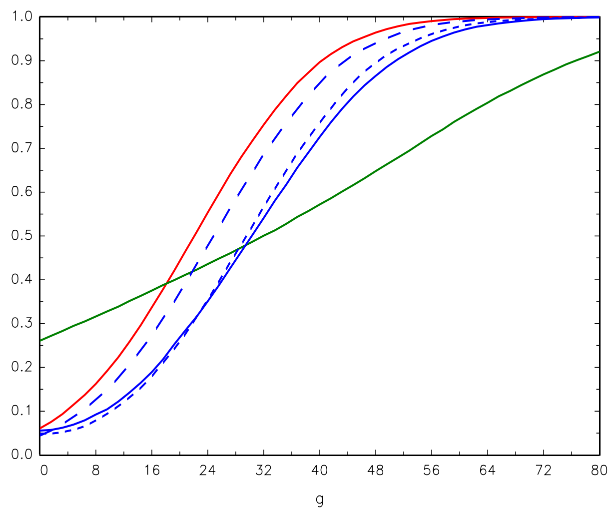
(c) $c = 25$



(d) $c = 50$

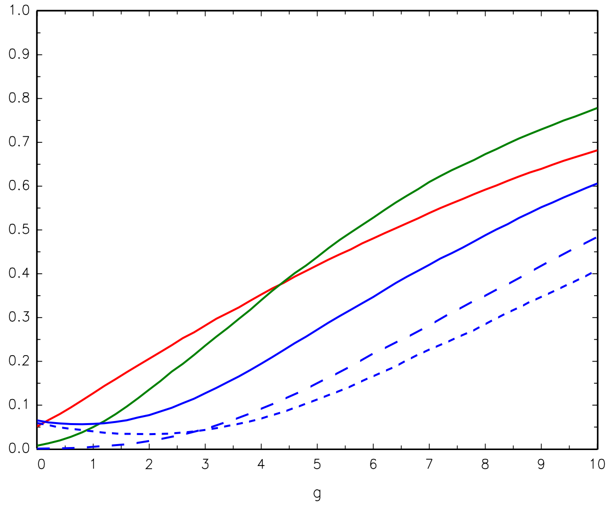


(e) $c = 100$

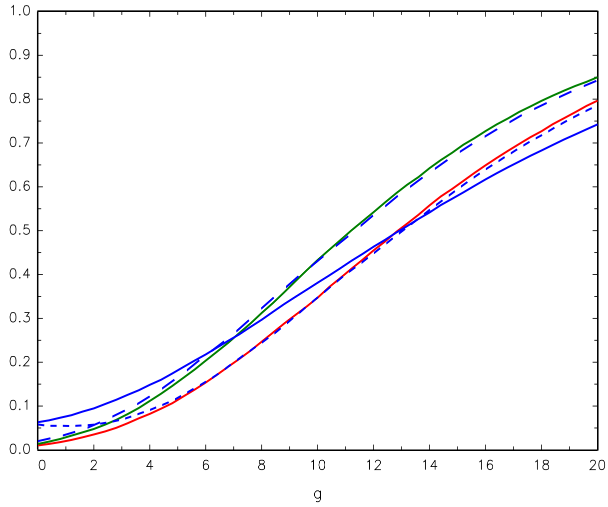


(f) $c = 200$

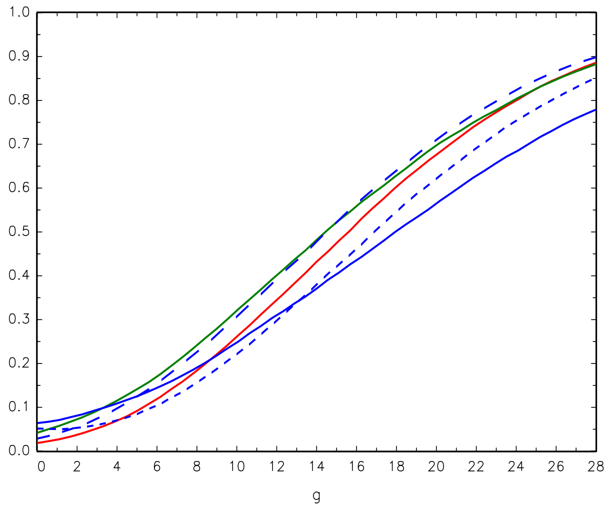
Figure 13. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.5$;
t_C^{wt}: —— , Q: —— , BD: —— , IVX_1: — — , IVX_2: - - -



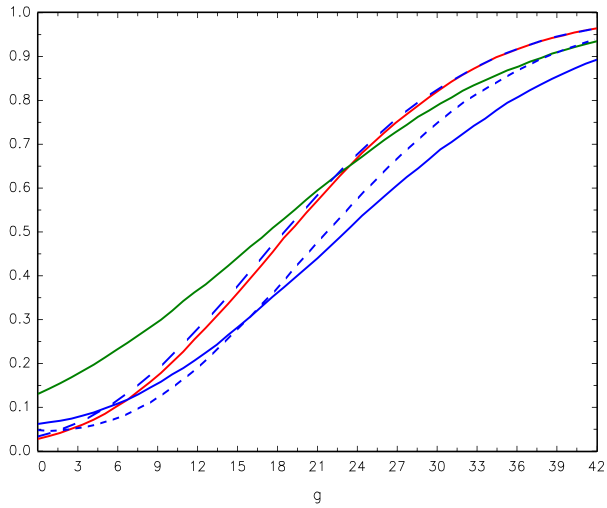
(a) $c = 0$



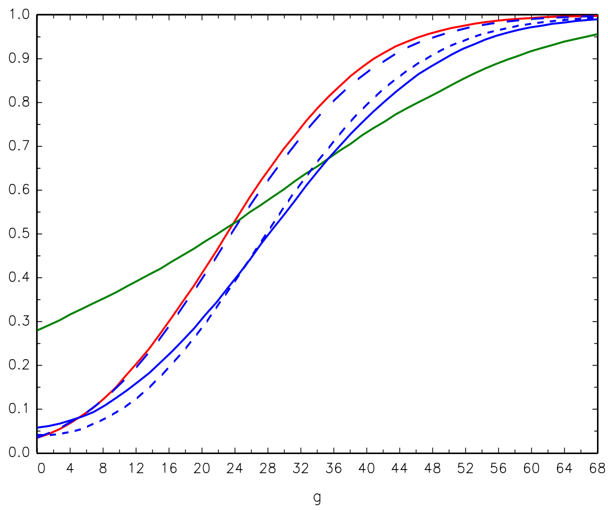
(b) $c = 10$



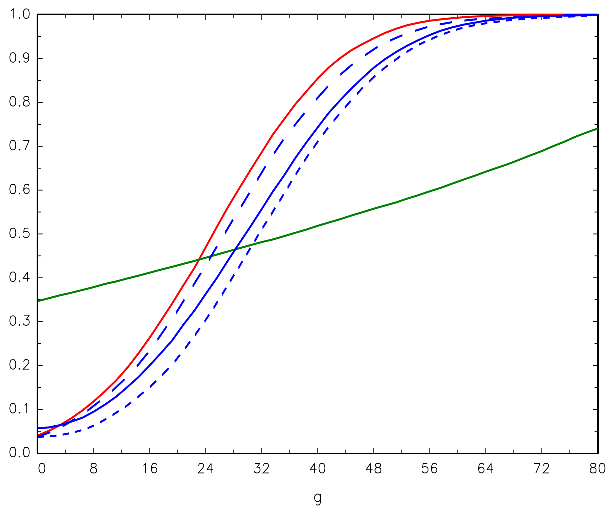
(c) $c = 25$



(d) $c = 50$



(e) $c = 100$



(f) $c = 200$

Figure 14. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.9$;
t_C^{wt}: —— , Q: —— , BD: —— , IVX_1: – – , IVX_2: - - -