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

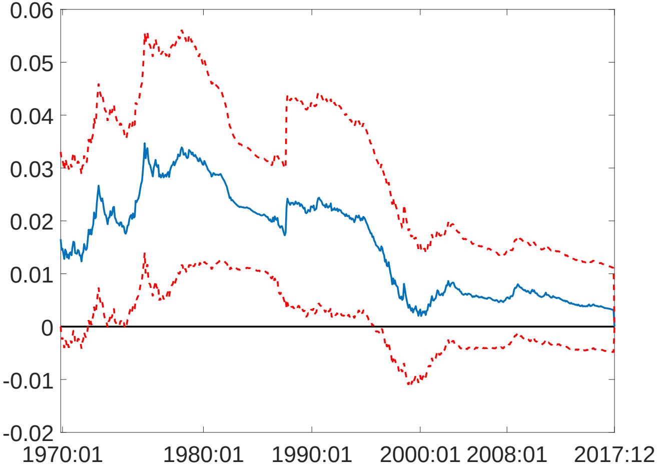
Robert Taylor, University of Essex

joint work with Dave Harvey and Steve Leybourne, University of Nottingham

ESRC Workshop, University of Essex, September 9th 2019

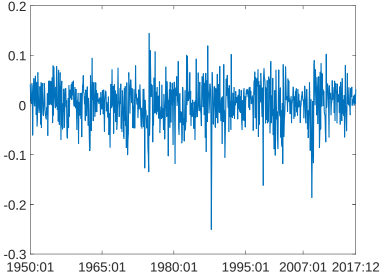
This research is funded by the ESRC under award number ES/R00496X/1.

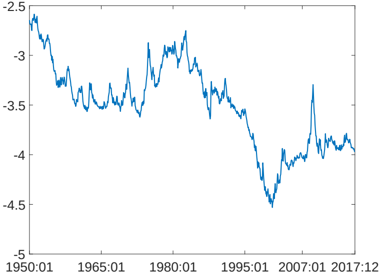
Is there any Predictability in the Equity Premium?



Dividend yield: Forward Recursive IV regression estimates and pointwise CIs, 1950-2017 (Goyal/Welch 2008 updated monthly data).

... what about the persistence of the predictor?





The equity premium looks very mean reverting etc (almost noise), but the dividend yield looks strongly persistent (usual ADF test has p-value of 0.41).

Outline

1. Background and Motivation
2. The Predictive Regression Model and Tests
3. Finite Sample Simulations I
4. Weighted Tests
5. Finite Sample Simulations II
6. Equity Premium Predictability
7. Concluding Remarks

Moving on to ...

The Basic Predictive Regression Set-up

Consider the predictive regression

$$y_t \quad = \quad \alpha_y + \beta\, x_{t-1} + \epsilon_{yt}$$

where

$$x_t \quad = \quad \phi\, x_{t-1} + \epsilon_{xt},$$

with $(\epsilon_{xt}, \epsilon_{yt})' \sim iid(0, \mathbf{\Sigma})$ where

$$\mathbf{\Sigma} = \mathrm{E}\left(\begin{pmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{pmatrix} \begin{pmatrix} \epsilon_{xt} & \epsilon_{yt} \end{pmatrix}\right) = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}.$$

Null hypothesis: x_{t-1} does not predict y_t, i.e.

$$H_0 : \quad \beta = 0\,.$$

Yet, even in this simplest setup...

Endogeneity and (high) Persistence

Should

- the shocks ϵ_{yt} and ϵ_{yt} correlate (so that $\rho_{xy} := \sigma_{xy}/\sigma_x\sigma_y \neq 0$; for the EP-DY data above this correlation is estimated to be $\hat{\rho}_{xy} = -0.98$), and
- the regressor x_t be autocorrelated,

one speaks of endogeneity. (A bit of a misnomer.)

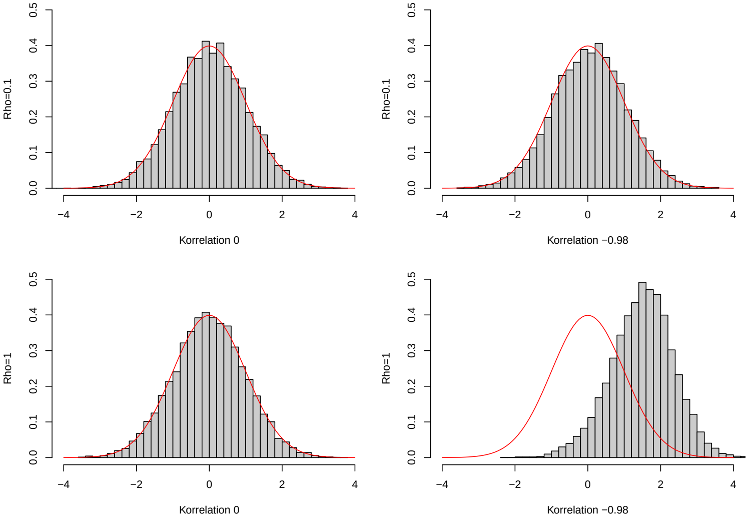
Under endogeneity and high persistence (near integration, $\phi = 1 - c/T$),

- the OLS estimator is 2nd order biased and
- the *t*-statistic has a non-normal limiting distribution.

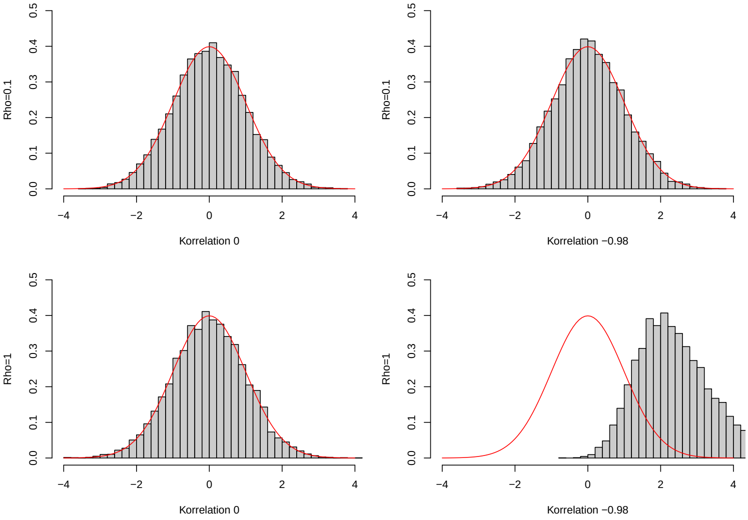
See Elliott/Stock (1994), Stambaugh (1999), Campbell/Yogo (2006) etc.

No problem when regressors are stationary or weakly persistent.

OLS t-statistics, $T = 305$



More trouble with variance breaks - volatility of both shocks 3 times higher in the first 20% of the sample



Popular Solutions in the Literature

If ϕ were known, one could employ GLS estimation. For unknown ϕ:

- Bayes methods - Elliott/Stock (1994)
- Bonferroni - Campbell/Yogo (2006), but see Phillips (2012)
- Restricted log-likelihood - Jansson/Moreira (2006), Chen/Deo (2009)
- Almost optimal tests - Elliott *et al.* (2015)
- Variable addition - Toda/Yamamoto (1995), Dolado/Lütkepohl (1996)
- Generic IV estimation - Breitung/Demetrescu (2015)
- IVX method of Kostakis *et al.* (2015)

Our Contribution to the Literature

- We develop new predictability tests in these circumstances based on simple regression t-ratios.
- The simplest test, optimal under Gaussianity for a weakly persistent and exogenous predictor, is based on the standard t-statistic.
- Where x_t is endogenous an optimal, but infeasible, test for predictability is based on the t-ratio on the lagged predictor when augmenting the basic predictive regression with the predictor's current period innovation, ϵ_{xt}.
- 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.

Our Contribution to the Literature

- In the near-unit root case, the limiting null distributions of these three statistics depend on both ρ_{xy} and c. We propose a feasible method for obtaining (conservative) asymptotic critical values and give response surfaces for these.
- 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 t-ratios discussed above, the weights obtained using the p-values from a unit root test on the predictor.
- Despite their simplicity, the weighted tests display very good finite sample size control and power across a range of persistence and endogeneity levels for the predictor, comparing favourably with the leading tests in the extant literature.

Moving on to ...

1. Background and Motivation
2. The Predictive Regression Model and Tests
3. Finite Sample Simulations I
4. Weighted Tests
5. Finite Sample Simulations II
6. Equity Premium Predictability
7. Concluding Remarks

The Predictive Regression Model

- We consider the predictive regression for y_t, the (excess) stock return in period t, by x_{t-1} a putative predictor variable at time $t-1$:

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

  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.
- The innovation vector $\epsilon_t := (\epsilon_{xt}, \epsilon_{yt})'$ is assumed for the presentation to be IID with finite fourth order moments and satisfying

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

- The paper shows how the methods discussed can be modified to allow for weak dependence and/or heteroskedasticity in ϵ_t.

Uncertain Persistence

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

Assumption 1

One of the following two conditions is assumed to hold:

1. ***Weakly persistent predictors: (W)*** *The autoregressive parameter ϕ in (2) is fixed and bounded away from unity, $|\phi| < 1$.*
2. ***Strongly persistent predictors: (S)*** *The autoregressive parameter ϕ in (2) is local-to-unity with $\phi := 1 - \frac{c}{T}$ where c is a fixed non-negative constant.*

Cholesky decomposition

- 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 $e_t := (e_{1t},\, e_{2t})' \sim IID\,(0, I_2)$.

- Using this representation, we can then re-write (1) as

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

which 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.

Null and Alternative Hypotheses

- We will test the null that y_t is not predictable by x_{t-1}, i.e. $H_0 : \beta = 0$ in (1), without the practitioner needing to know which of Assumption S or Assumption W holds in (2).
- The alternative hypothesis is that y_t is predictable by x_{t-1}, ie. $\beta \neq 0$.
- Predictive regressions for stock returns typically exhibit a small R^2 and low signal-to-noise ratios so departures from the null are likely to be small. We therefore focus on local alternatives such that the slope parameter β is local-to-zero.
- 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.

Predictability Tests - An Infeasible Test

- Consider the 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).
- 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).
- Denote the infeasible t-statistic as t^{inf}. This has a standard normal limiting distribution under H_0, under either Assumption S or Assumption W. Moreover, under Gaussianity this would deliver an efficient test (among α_y, α_x invariant tests) whenever $\rho_{xy} \neq 0$.
- Including ϵ_{xt} as a regressor reduces the error variance from σ_y^2 to $\sigma_y^2(1 - \rho_{xy}^2)$; ie, with knowledge of ϕ we can subtract off the part of the innovation to returns that is correlated with the innovation to the predictor variable, delivering a more powerful test.

Feasible Tests I - a standard t-statistic

- The standard t-statistic based on (1) is effectively based on an OLS regression that omits z_{xt} from (5). Denote this statistic as t.
- 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 (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 $\rho_{xy} \neq 0$, its limit null distribution depends on ρ_{xy} and c, and is expected to be highly inefficient in this strongly persistent case, due to the lack of any proxy for ϵ_{xt} in the underlying regression.

Feasible Tests II - a variant t-statistic

- We also consider a variant of the standard t-statistic that might be considered more appropriate for 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. More natural 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

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

  which we denote by t'.
- Under Assumption S the asymptotic null distribution of t' will depend on ρ_{xy} and c. Under Assumption W, the limiting null distribution of t' will depend on the (unknown) distribution of s_1, hence t' is only designed for use in the strongly persistent world (in contrast to t).

Feasible Tests III - A t-stat based on a proxy measure for ϵ_{xt}

- Can we obtain a proxy measure for ϵ_{xt}? There are a number of ways this could be done. Campbell and Yogo (2006) use a Bonferroni-based method. We consider an approach based on including a covariate z_{xt} in (5) to act as a direct proxy for ϵ_{xt} in (4).
- Under Assumption S a proxy for ϵ_{xt} can be obtained by assuming a particular value for c, say $\bar{c}$. Then construct $z_{xt} = x_t - (1 - \bar{c}T^{-1})x_{t-1}$ (assuming $\alpha_x = 0$ for simplicity). If $\bar{c} = c$, then $z_{xt} = \epsilon_{xt}$ and we obtain the asymptotically standard normal and efficient test, t^{inf}. But if $\bar{c} \neq c$ the critical values for this test will depend on both ρ_{xy} and c, and it will no longer be efficient, with power being a (decreasing) function of the distance $|c - \bar{c}|$. Recall that c cannot be consistently estimated.
- An obvious proxy for ϵ_{xt} is given by setting z_{xt} equal to the OLS estimate, $\hat{\epsilon}_{xt}$ say, obtained from an OLS regression of Δx_t on a constant and x_{t-1}. But this residual is exact orthogonal to x_{t-1} and so the resulting t test delivers identical inference to t.

Feasible Tests III - A t-stat based on a proxy measure for ϵ_{xt}

- An alternative proxy which takes account of a strongly persistent autoregressive structure in estimating the intercept term α_x, is to use the GLS-type estimator of α_x as in t'. So set $z_{xt} = \tilde{\epsilon}_{xt}$ in (5) the residuals from regressing Δx_t on $(x_{t-1} - x_1)$.
- Unlike $\hat{\epsilon}_{xt}$, $\tilde{\epsilon}_{xt}$ is not orthogonal to x_{t-1} which raises the potential for $\tilde{\epsilon}_{xt}$ to proxy ϵ_{xt} in the strongly persistent case. The resulting t-statistic is denoted t^*.
- Under Assumption S, the limiting null distribution of t^* depends on ρ_{xy} and c in the case where x_t is strongly persistent (Assumption S). Under Assumption W s_t and s_1 are both of $O_p(1)$, and the asymptotic distribution t^*, like t', depends on the distribution of s_1.

Asymptotic Critical Values I

- Under Assumption S the limiting null critical values of t, t',and t^* depend on the unknown parameters ρ_{xy} and c.
- Although ρ_{xy} can be consistently estimated, c cannot. We therefore adopt a scheme for simulating critical values designed to yield asymptotically conservative tests.
- To illustrate, consider t and denote its null limit distribution under Assumption S by $\mathcal{S}(0, \rho_{xy}, c)$. For expository purposes we will focus attention on upper-tail critical values 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 next.

Asymptotic Critical Values II

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 $cv_\alpha(\rho_{xy}) := \max_{c \in [0, c_{\max}]} cv_\alpha(\rho_{xy}, c)$.

Asymptotic Critical Values III

- 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.
- Direct simulation methods were used across values of $\rho_{xy} \in [-0.950, -0.925, -0.900, ..., 0.900]$ with $c_{\max} = 50$, evaluating the critical values for $c \in [0, 2, 4, ..., 50]$ for $\alpha = 0.05$ 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 $c_{\max}$; eg, for $\rho_{xy} = -0.9$, it is obtained at $c = 0$ for both values of α.
- For a given value of ρ_{xy}, we give a response surface from 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}.

Asymptotic Critical Values IV

- In practice, substitute ρ_{xy} by a consistent estimate in the fitted response surface.
- To that end, we use

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

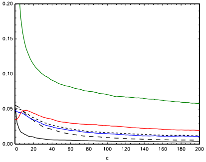
  $\hat{\epsilon}_{yt}$ are the OLS residuals from regressing y_t on a constant and x_{t-1}, and $\hat{\epsilon}_{xt}$ are the OLS residuals from regressing Δx_t on a constant and x_{t-1}. This estimator is consistent under either Assumption S or Assumption W.
- Tests based on comparison of t, t' and t^* with their asymptotically conservative critical values will be denoted by t_C, t'_C and t^*_C, respectively.

Moving on to ...

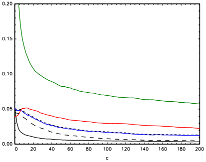
Monte Carlo Design

- Compare the finite sample properties of the t_C, t'_C and t^*_C tests for $T = 200$ (qualitatively similar results for other sample sizes).
- 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.
- 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 pure random walk and a pure white noise process.
- We conduct upper-tailed tests at the nominal 0.05 level, using the asymptotic conservative critical values obtained as described just now. All based on 20,000 Monte Carlo replications.

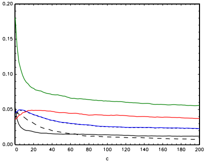
Back



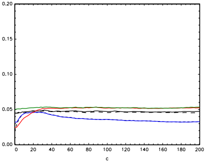
(a) $\rho_{xy} = -0.95$



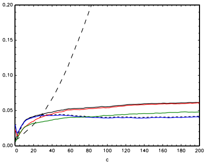
(b) $\rho_{xy} = -0.9$



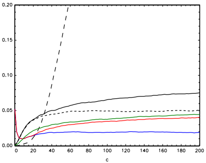
(c) $\rho_{xy} = -0.5$



(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: —— , t_C: —— , t'_C: - - - - , t^*_C: — — , t^w_C: —— , t^{wt}_C: ——

Empirical Size Properties I

- Figure 1 - sizes of t_C, t'_C and t^*_C across c for different values of ρ_{xy}. As a size benchmark also include 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.
- Panel (a) $\rho_{xy} = -0.95$:
  - 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.
  - t_C has near-correct empirical size with $c = 0$, but undersized for $c > 0$, 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.
  - Sizes of t'_C and t^*_C 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.
  - t'_C holds on to size better for larger values of c than t^*_C which may have implications for the relative power of the two in this region.

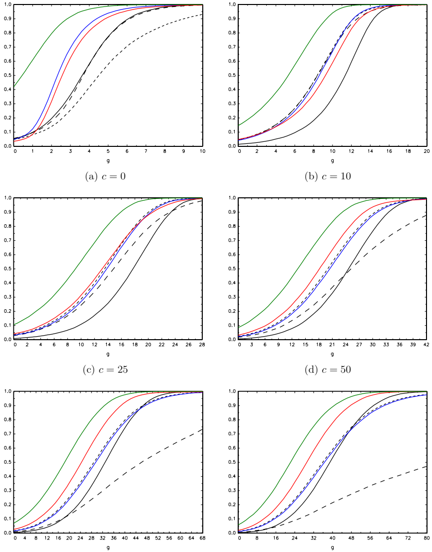
Empirical Size Properties II

- Panel (b), $\rho_{xy} = -0.90$: very similar to those for $\rho_{xy} = -0.95$.
- Panel (c), $\rho_{xy} = -0.5$: t_N remains over-sized, the other three tests tend to be less under-sized for large c, particularly t'_C.
- Panel (d), $\rho_{xy} = 0$: all four tests display decent empirical size across c. As expected, sizes of t_C and t_N are similar.
- Panels (e) and (f): $\rho_{xy} = 0.5, 0.95$: all 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. 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, for negative ρ_{xy}, the best finite sample size control is given by t'_C for larger c, with little to choose between t'_C and t^*_C for smaller c. For positive ρ_{xy}, t^*_C is best avoided unless c is small.

Empirical Power Properties I

- Next we simulate powers across g (≥ 0), with $\beta = g/200$, for different values of ρ_{xy} and c.
- Figure 2, $\rho_{xy} = -0.95$:
  - Note here the powers of t_N are fairly meaningless outside of panel (f) where $c = 200$, due to its over-size discussed above.
  - For $c = 0$, 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' have similar power levels both substantially higher than t_C. Same largely true for $c = 25$.
  - For $c = 50$, 100 and 200, t_C^* falls well behind t_C'. For $c = 200$, t_N has correct size and, unsurprisingly given the under-sizing seen in the other tests, the most power.
  - On the basis of these results, t_C' and t_C^* give the best performance across small to moderate values of c, yet we would hold a distinct preference for t_C' for the larger values of c.

Back



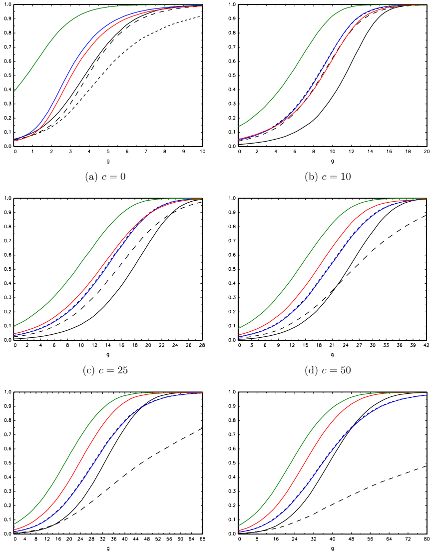
(a) $c = 0$ (b) $c = 10$

(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

Back

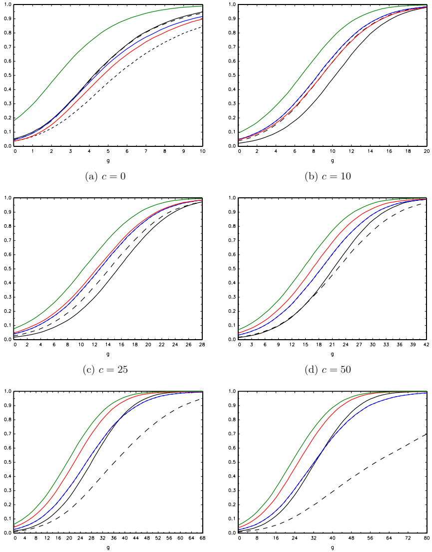


(a) $c = 0$ (b) $c = 10$

(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: —— , t_C: —— , t'_C: - - - - , t^*_C: — — , t^w_C: —— , t^{wt}_C: ——

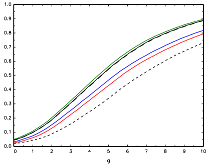


(a) $c = 0$ (b) $c = 10$

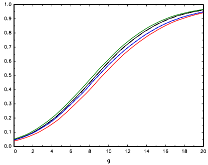
(c) $c = 25$ (d) $c = 50$

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

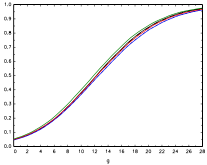
Figure 4. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.5$;
t_N: green solid, t_C: black solid, t'_C: black short-dashed, t^*_C: black long-dashed, t^w_C: blue solid, t^{wt}_C: red solid



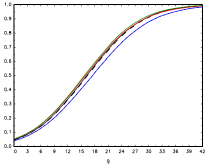
(a) $c = 0$



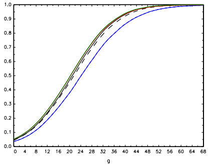
(b) $c = 10$



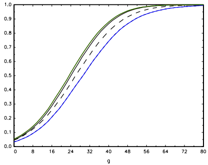
(c) $c = 25$



(d) $c = 50$

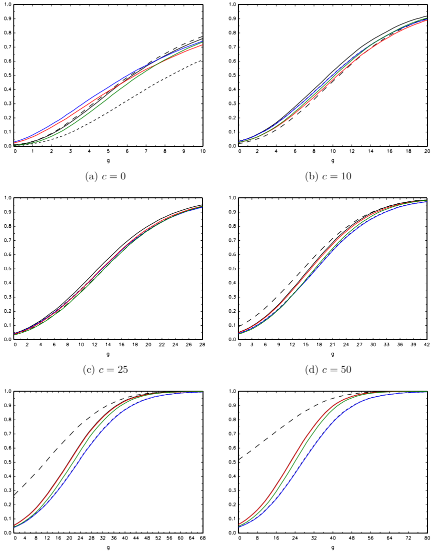


(e) $c = 100$



(f) $c = 200$

Figure 5. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0$;
t_N: 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

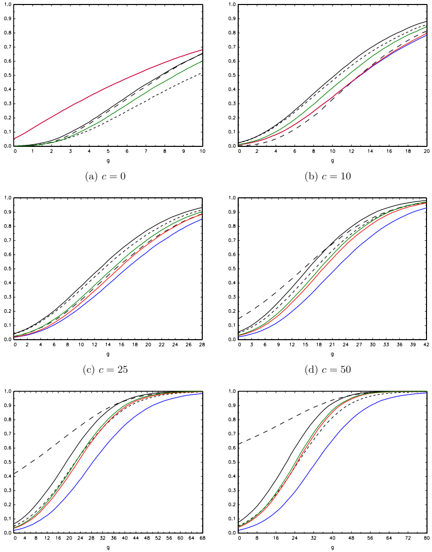


(a) $c = 0$ (b) $c = 10$

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

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

Figure 6. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.5$;
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) $c = 0$ (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: ——, t_C: ——, t'_C: - - - -, t^*_C: — —, t^w_C: ——, t^{wt}_C: ——

Empirical Power Properties II

- For $\rho_{xy} = -0.9$ and $\rho_{xy} = -0.5$ (Figures 3 and 4), we draw similar conclusions.
- 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.

Moving on to ...

1. Background and Motivation
2. The Predictive Regression Model and Tests
3. Finite Sample Simulations I
4. Weighted Tests
5. Finite Sample Simulations II
6. Equity Premium Predictability
7. Concluding Remarks

Weighted Tests I

- Given the foregoing simulation results 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.

Weighted Tests II

- An obvious way to do this is to use a simple weighting based on the p-value from a unit root test applied to x_t with 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.
- Consider a generic unit root statistic, UR, and denote by p^{UR} the (asymptotic) p-value associated with UR.
- Based on p^{UR} we can then define a weighted statistic, denoted t^w, as

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

  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'.

Weighted Tests III

- We consider a further weighted test designed to improve power under low persistence.
- Under Assumption W, the t_N test is valid with an attractive power profile; see $c = 200$ results for $\phi = 0$. For negative ρ_{xy}, the power of t_C test is adversely affected because t has to be be compared with conservative critical values dominated by the behaviour of t at $c = 0$.
- Additional power could be obtained by integrating t into the weighted statistic when c is very large. To that end,

$$t^{wt} := t^{w}\mathbb{I}(p^{UR} > \tau) + t\mathbb{I}(p^{UR} \leq \tau), \ \ \tau > 0$$

- If the user-specified cut-off value τ is chosen suitably small, then for small c, t^{wt} will typically be t^{w} rather than t, so t cannot inflate the critical values of t^{wt} too 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.

Weighted Tests IV

- Under Assumption S, 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.
- Under Assumption W, t^{wt} and t are asymptotically equivalent, by virtue of that fact that τ is a positive constant.
- So again we need to obtain simulated asymptotic 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.

Asymptotic Critical Values

- For UR we chose the familiar local-GLS detrended Dickey-Fuller statistic of Elliott *et al.* (1996) for $\bar{c} = 0$, denoted DF; ie, the OLS t-statistic for $\phi = 0$ in the regression model

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

- DF satisfies the necessary regularity conditions for UR, and is simple to compute. We found that it led to weighted tests with superior finite sample properties to other common unit root tests. Based on experimentation, we set $\lambda = 2$ and $\tau = 0.01$.
- We used direct simulation to obtain limiting p-values for DF and give a self-contained response surface, but a valid bootstrap could also be used.
- Conservative asymptotic critical values for the t^w and t^{wt} statistics are calculated in essentially the same manner as for t, t' and t^*. A response surface is given in the paper. Denote the tests based on comparison of t^w and t^{wt} with these critical values by t^w_C and t^{wt}_C.

Moving on to ...

Empirical Size

- Figure 1 Striking similarity between the size properties of t_C^w and t'_C. Only for $\rho_{xy} = 0.9$ do their sizes appear to differ in any way, with t_C^w being more under-sized across c.
- For positive ρ_{xy}, t_C^w does not exhibit the large over-sizing outside of small c associated with t_C^*. Why? 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.
- t_C^{wt} has accurate size across c for negative values of ρ_{xy}, establishing it as easily the most reliably sized test in those cases. Elsewhere, some under-sizing for small c, although as this is generally a feature of all of the tests, t_C^{wt} is still very competitive. Comparing t_C^w and t_C^{wt}, for 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 $c = 0$.

Empirical Power I

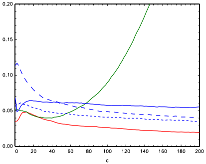
- $\rho_{xy} = -0.95$: Figure 2
  - For $c = 0$, t_C^w and t_C^{wt} emerge as easily the most powerful tests (ignore t_N as grossly over-sized here) with similar power, both well exceeding the power of the third placed tests t_C and t^*.
  - For $c = 10$ and $c = 25$, t'_C, t^*_C, t_C^w and t_C^{wt} show broadly similar power profiles.
  - For $c = 50$ t_C^{wt} emerges as most powerful, with t_C^w tying for second place with t'_C. This because t^w is now placing most weight on t', while t^{wt} is now very regularly switching into t. Same rankings for $c = 100$ and 200, but the power advantage of t_C^{wt} over t_C^w increases with c. Notice for $c = 200$, t_C^{wt} has power approaching the levels associated with t_N (which is correctly sized here).
- Same observations regarding power ranking also apply for $\rho_{xy} = -0.9$ and $\rho_{xy} = -0.5$ Figure 3 and Figure 4.

Empirical Power II

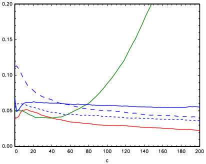
- For $\rho_{xy} = 0$ Figure 5 neither t_C^w nor t_C^{wt} dominates the best of the other tests, but the power differences involved are very modest.
- For $\rho_{xy} = 0.5$ Figure 6, t_C^{wt} is arguably the best performing test overall, albeit by a small margin. For the larger c values, we see that t_C^{wt} again outperforms t_C^w.
- 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 earlier. For large c, the dominance of t_C^{wt} over t_C^w is quite evident.
- Summing up, 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.

Comparison with Extant Tests

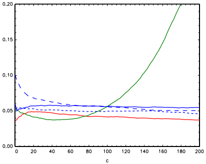
- We compare t_C^{wt} with: Campbell and Yogo's Q test (one-sided, upper tail), the (two-sided) instrumental variable test of Breitung and Demetrescu (2015) using sine and fractional instruments, denoted BD; and the IVX test of Kostakis *et al.* (2015), with IVX_1 and IVX_2 denoting one-sided (upper tail) and two-sided tests, respectively.
- All tests were run at the 5% level and 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).



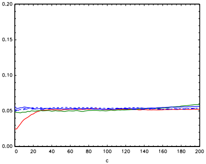
(a) $\rho_{xy} = -0.95$



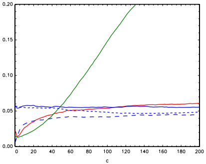
(b) $\rho_{xy} = -0.9$



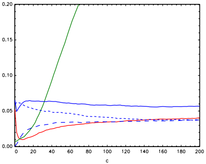
(c) $\rho_{xy} = -0.5$



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



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

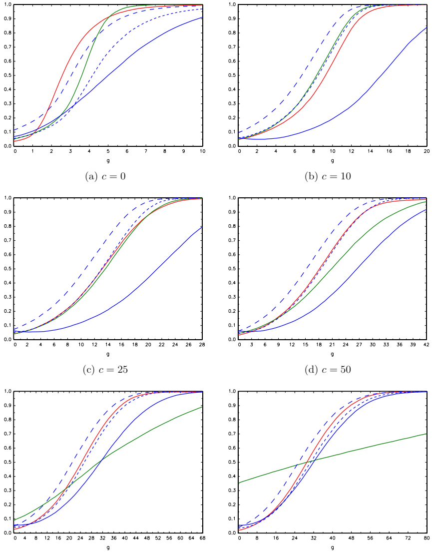


(f) $\rho_{xy} = 0.9$

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

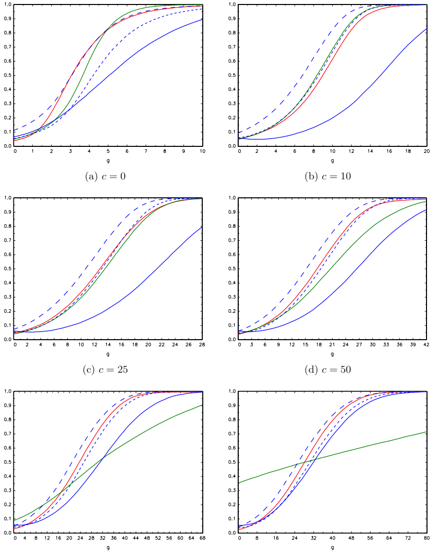
Empirical Size

- The standout feature in 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 seen 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 it avoids the issue of over-sizing seen with other one-sided tests.



(a) $c = 0$ (b) $c = 10$ (c) $c = 25$ (d) $c = 50$ (e) $c = 100$ (f) $c = 200$

Figure 9. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.95$;
t_C^{wt}: —, Q: —, BD: —, IVX_1: – –, IVX_2: - - -



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

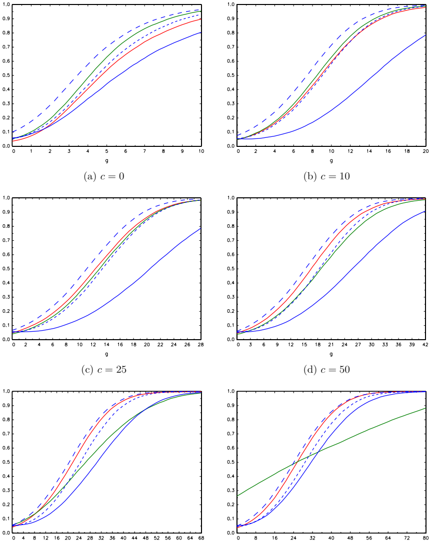
(c) $c = 25$ (d) $c = 50$

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

Figure 10. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.9$;
t_C^{wt}: —, Q: —, BD: —, IVX_1: – –, IVX_2: - - -

Empirical Power I

- $\rho_{xy} = -0.95$ Figure 9:
  - For $c = 0$ (ignoring IVX_1 due to its significant over-size discussed above) 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.
  - 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.
- Results for $\rho_{xy} = -0.9$ Figure 10 are very similar to those for $\rho_{xy} = -0.95$.



(a) $c = 0$

(b) $c = 10$

(c) $c = 25$

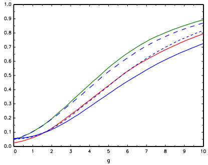
(d) $c = 50$

(e) $c = 100$

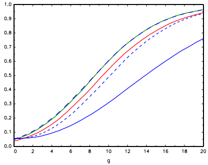
(f) $c = 200$

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: - - -

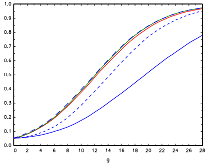
Back



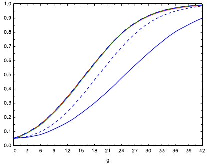
(a) $c = 0$



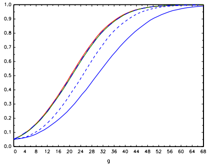
(b) $c = 10$



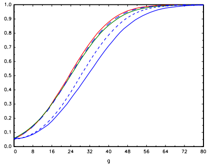
(c) $c = 25$



(d) $c = 50$



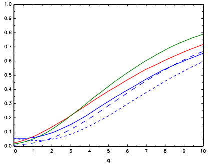
(e) $c = 100$



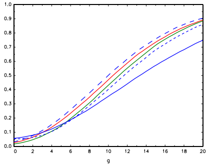
(f) $c = 200$

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: - - -

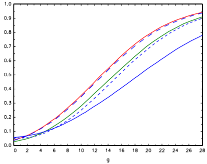
Back



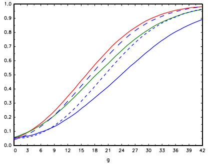
(a) $c = 0$



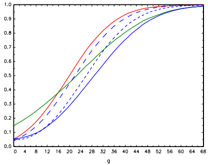
(b) $c = 10$



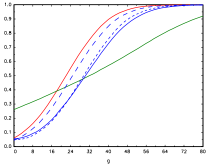
(c) $c = 25$



(d) $c = 50$



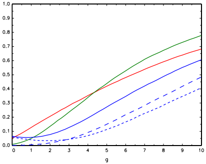
(e) $c = 100$



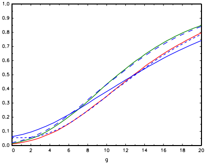
(f) $c = 200$

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: - - -

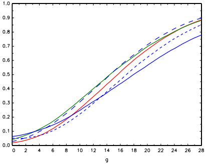
Back



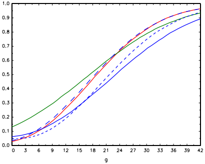
(a) $c = 0$



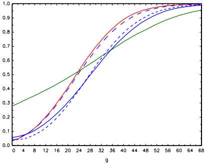
(b) $c = 10$



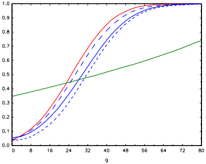
(c) $c = 25$



(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: - - -

Empirical Power II

- For $\rho_{xy} = -0.5$ 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.
- For $\rho_{xy} = 0.5$ Figure 13 t_C^{wt} arguably 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 $c = 0$.

Summary

- 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 would have the best finite sample size and power properties across the full constellation of settings for the persistence level in the predictive regressor and the correlation coefficient between the innovations in the model we have considered.
- 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.

Moving on to ...

Data

- Monthly U.S. equity series analysed in Welch and Goyal (2008), using updated data for the period 1970:1-2017:12 ($T = 576$) available at http://www.hec.unil.ch/agoyal/.
- Dependent variable, y_t, is S&P 500 value-weighted log excess return.
- We consider thirteen putative predictor variables for x_t: 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 book-to market value ratio. Details in Welch and Goyal (2008).

Test Procedures Applied

- Apply t_C, t'_C, t^*_C, t^w_C and t^{wt}_C, employing $UR = DF$, $\lambda = 2$ and $\tau = 0.01$.
- Lagged difference terms in Δx_t were 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 $p_{\max} = 12$.
- 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 $p_{\max} + 1$.

Results I

- Of these series, we report the five where at least one of the nine tests considered yields a rejection at the 0.10 level: the dividend-price ratio (d/p), dividend yield (d/y), default return spread (dfr), inflation rate (inf) and the stock variance ($svar$).
- We also report the values of $\hat{\rho}_{xy}$ and p^{DF} for these five series.

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.

Results II

- For the strongly persistent predictors d/p and d/y, 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_C^{wt}, delivers a rejection - exactly the sort of environment where our simulation results showed t_C^{wt} to be more powerful than t_C, t'_C and t^*_C.

Results II

- The remaining predictors dfr, inf and $svar$ do not appear to be strongly persistent (in each case p^{DF} is close to zero).
- For these series t^w places virtually all weight on t'. For dfr and $svar$, the low persistence prompts t^{wt} to switch into t. For dfr, this switch turns a 0.10-level non-rejection by t_C^w into a 0.05 level rejection by t_C^{wt}. This accords well with our simulation evidence which showed t_C^{wt} to be more powerful than t_C^w under weak persistence.
- All of the tests 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 them. Maybe due to the bad oversize of Q with low-persistence predictors?
- Focussing on the outcomes of t_C^{wt}, 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.

Moving on to ...

1. Background and Motivation
2. The Predictive Regression Model and Tests
3. Finite Sample Simulations I
4. Weighted Tests
5. Finite Sample Simulations II
6. Equity Premium Predictability
7. Concluding Remarks

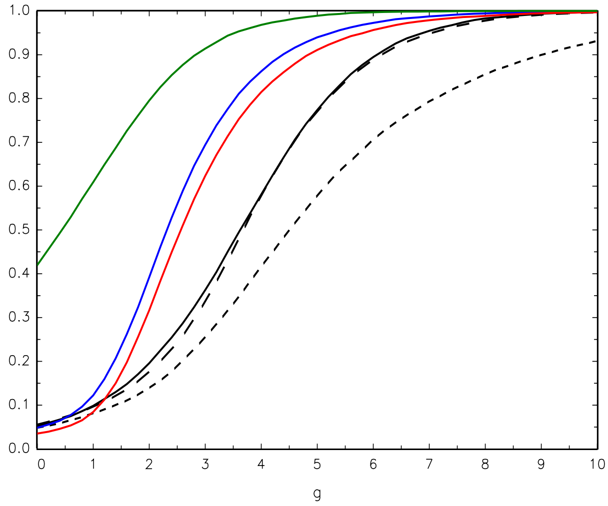
Summing up

- Develop new and easy to implement tests for predictability based on simple to compute regression t-ratios
- A weighted test, based on the two best performing of these t tests, is proposed designed to select the test with the better properties depending on the strength of persistence in the putative predictor
- Weights obtained according to the p-values from a standard Dickey-Fuller-type unit root test on the predictor
- A further modification combines the weighted test with the standard t-ratio reverting to the latter under very low persistence. Compares very favourably with the leading tests in the literature.
- In an empirical application to US stock returns, we uncover evidence of predictability using our new tests for a number of predictors with significantly differing persistence and endogeneity characteristics.

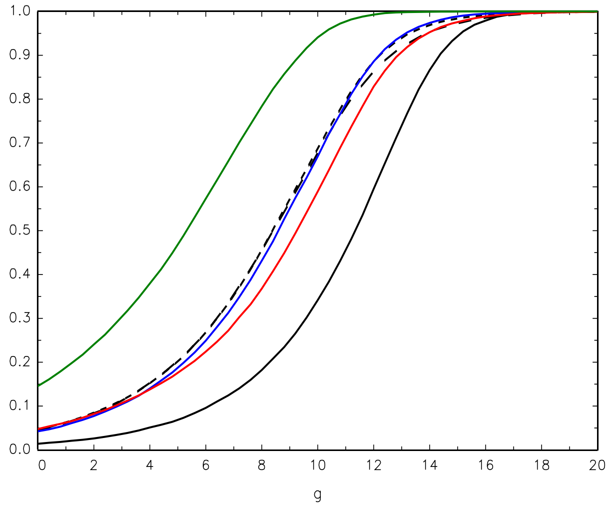
Welch and Goyal data

Detailed description of the variables used can be found on Amit Goyal's web page.

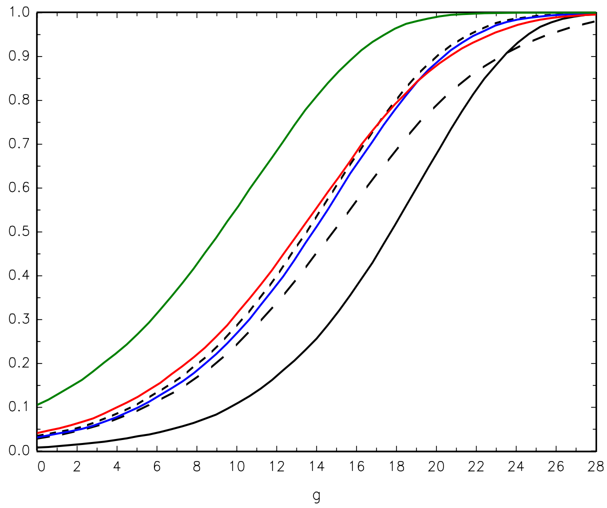
- The dependent variable, y_t, is the equity premium, EP_t which corresponds to the total rate of return of the S&P 500 index minus a short-term (risk-free) interest rate.
- DP_t is the dividend price ratio - difference between the log of dividends and the log of prices.
- DY_t is the dividend yield - difference between the log of dividends and the log of lagged prices.
- E/P_t is the earnings price ratio - difference between log of earnings and log of prices.
- DE_t is the dividend payout ratio - difference between log of dividends and log of earnings.
- $RVOL_t$ is the equity risk premium volatility.
- $SVAR_t$ is the stock variance computed as sum of squared daily returns on the S&P 500.
- BM_t is the book to market ratio - ratio of book value to market value for the Dow Jones Industrial Average.
- $NTIS_t$ is the net equity expansion - ratio of twelve-month moving sums of net issues by NYSE listed stocks divided by the total market capitalization of NYSE stocks.
- tbl_t is the treasury bill rate.
- lty_t is the long-term government bond yield.
- ltr_t is the long-term government bond rate of return.
- tms_t is the term spread - difference between the long-term yield on government bonds and the treasury bill rate.
- dfy_t is the default yield spread - difference between BAA- and AAA- rated corporate bond yields.
- dfr_t is the default return spread - difference between the return on long-term corporate bonds and returns on the long-term government bonds.
- $INFL_t$ is inflation - the consumer price index (all urban consumers).



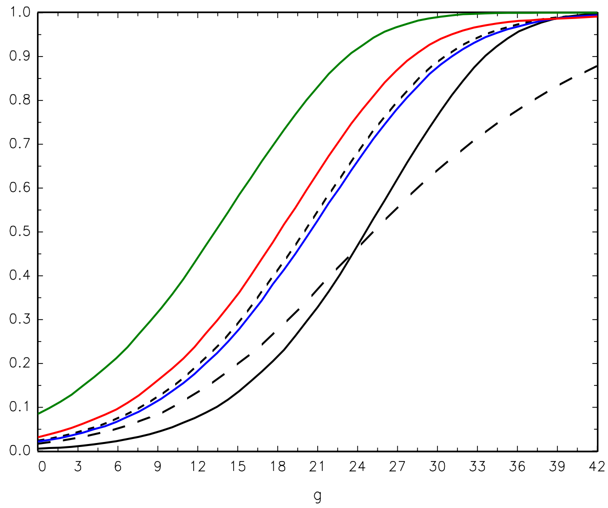
(a) $c = 0$



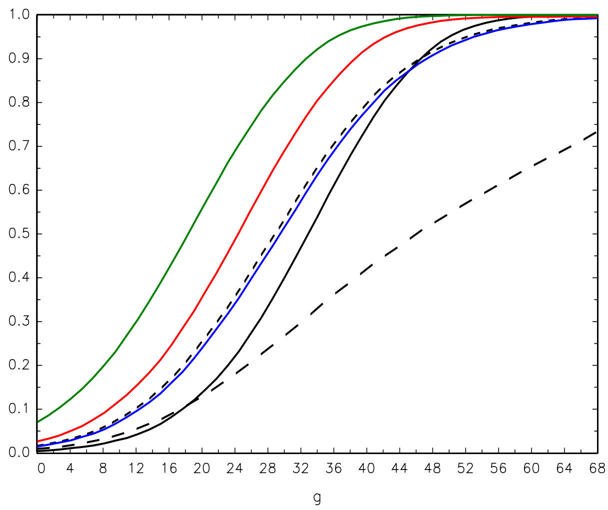
(b) $c = 10$



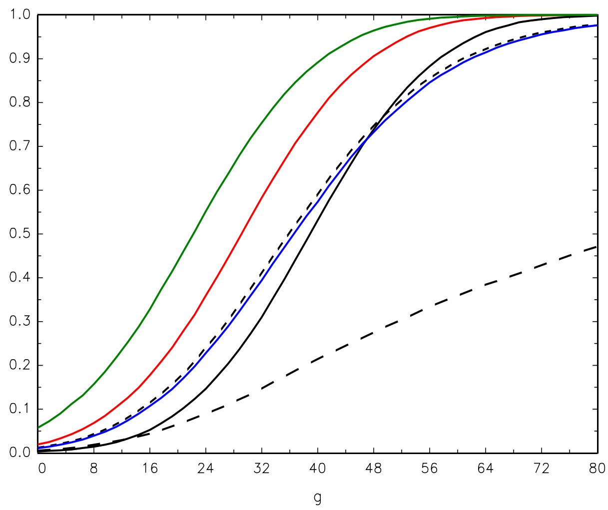
(c) $c = 25$



(d) $c = 50$

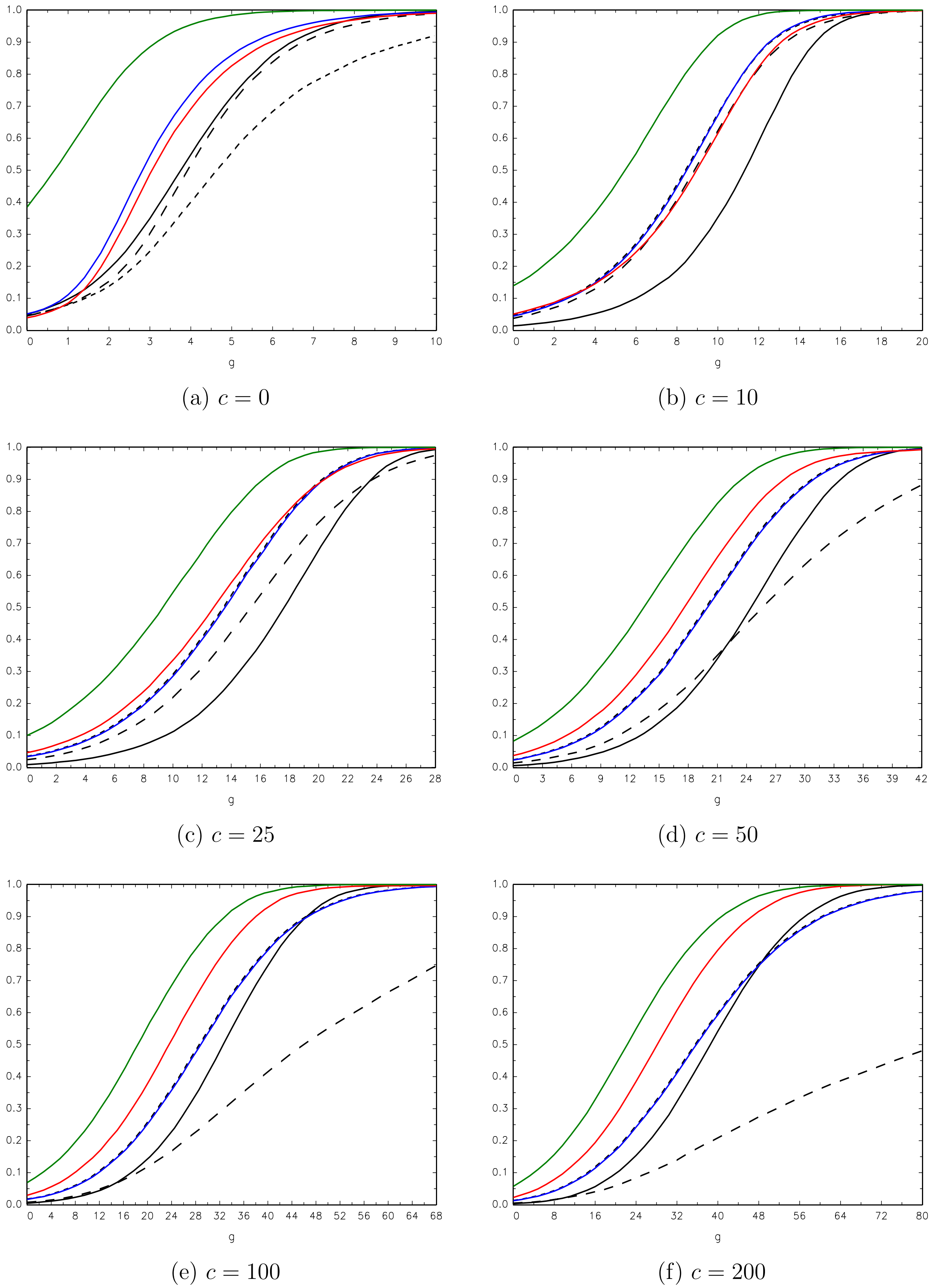


(e) $c = 100$



(f) $c = 200$

Figure 2. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.95$; t_N: —— , t_C: —— , t'_C: - - - , t^*_C: — — , t^w_C: —— , t^{wt}_C: ——

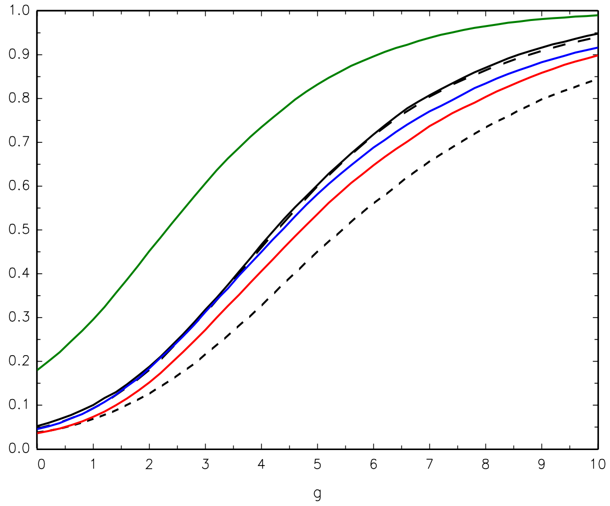


(a) $c = 0$ (b) $c = 10$

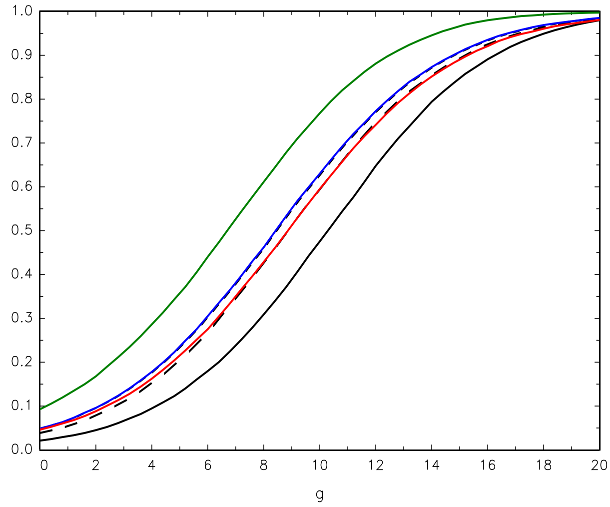
(c) $c = 25$ (d) $c = 50$

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

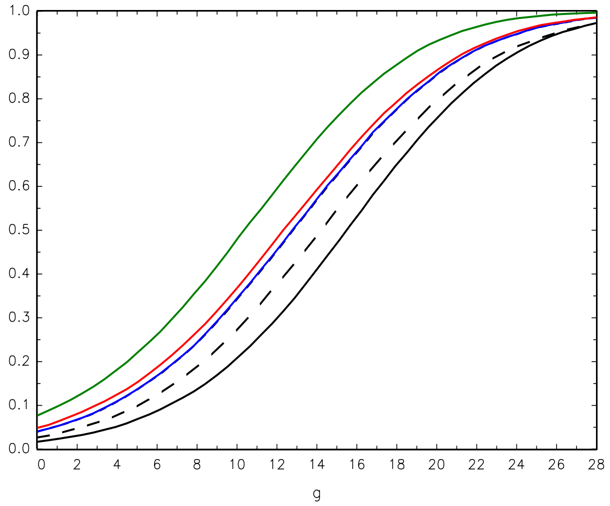
Figure 3. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.9$;
t_N: green solid line, t_C: black solid line, t'_C: short-dashed line, t^*_C: long-dashed line, t^w_C: blue solid line, t^{wt}_C: red solid line



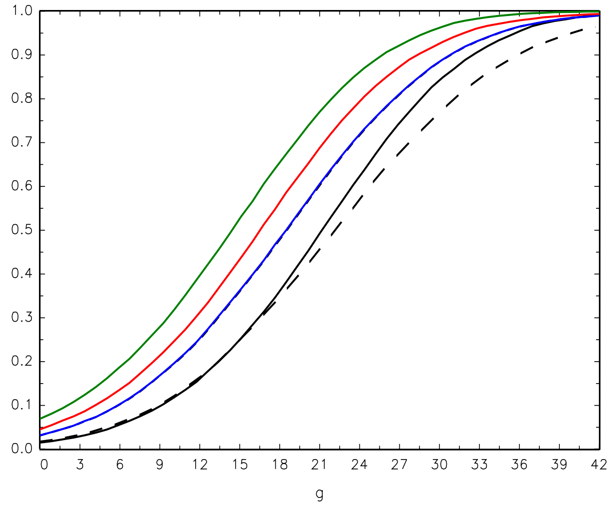
(a) $c = 0$



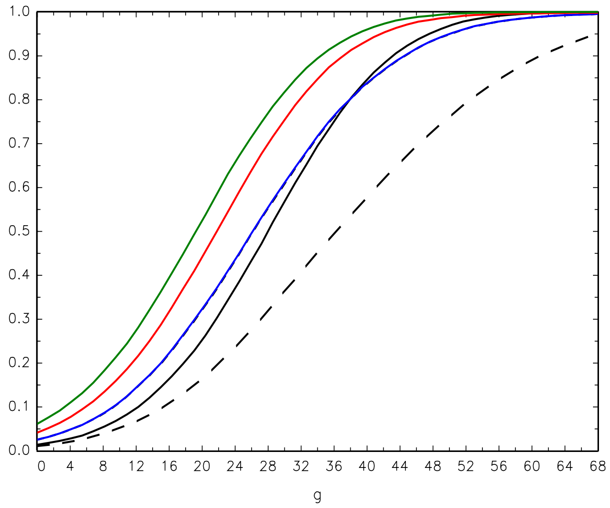
(b) $c = 10$



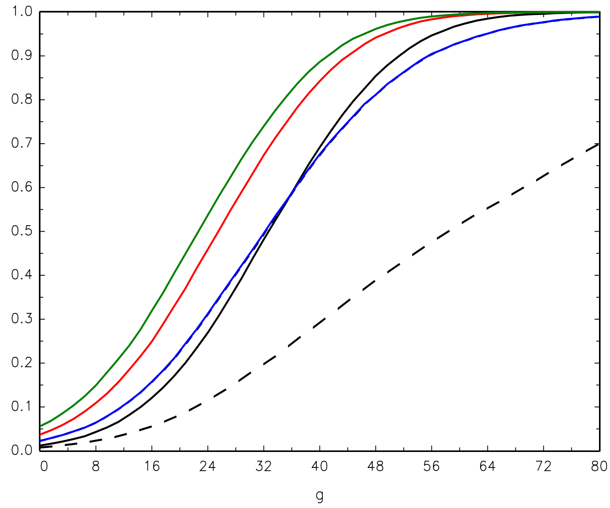
(c) $c = 25$



(d) $c = 50$

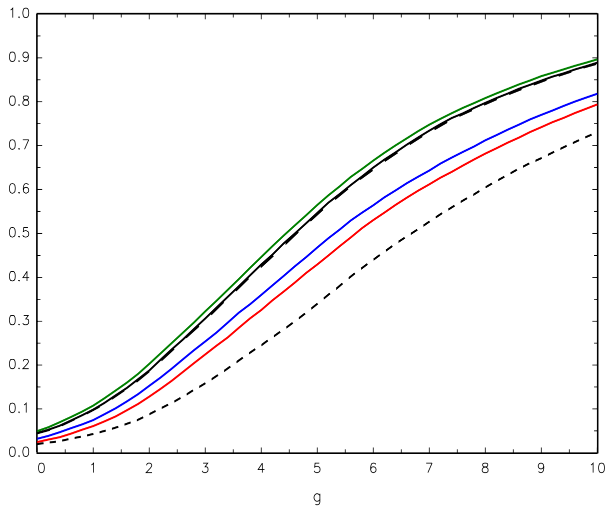


(e) $c = 100$

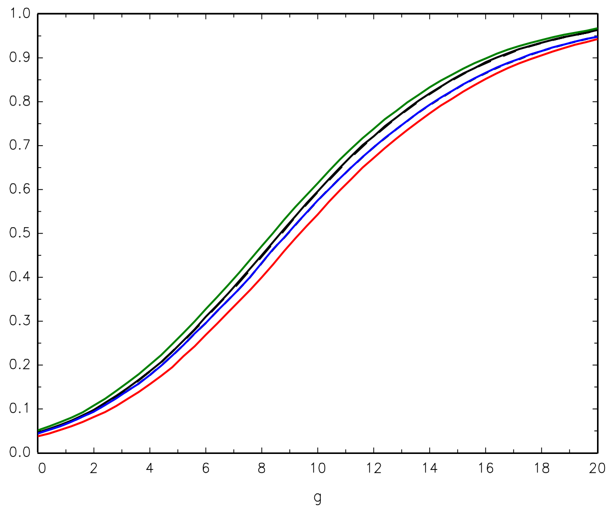


(f) $c = 200$

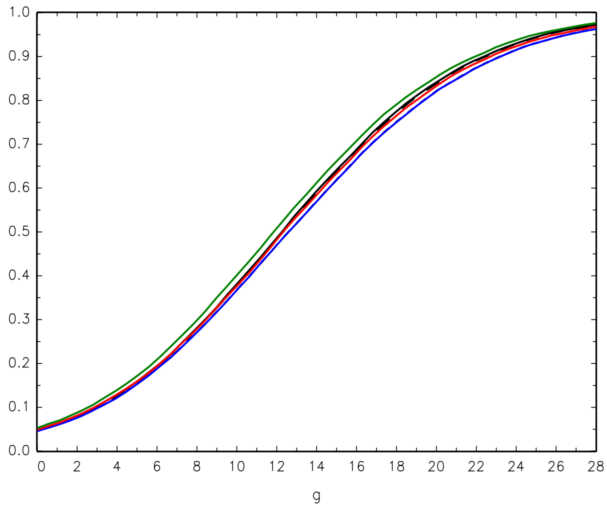
Figure 4. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.5$; t_N: ——, t_C: ——, t'_C: - - -, t^*_C: — —, t^w_C: ——, t^{wt}_C: ——



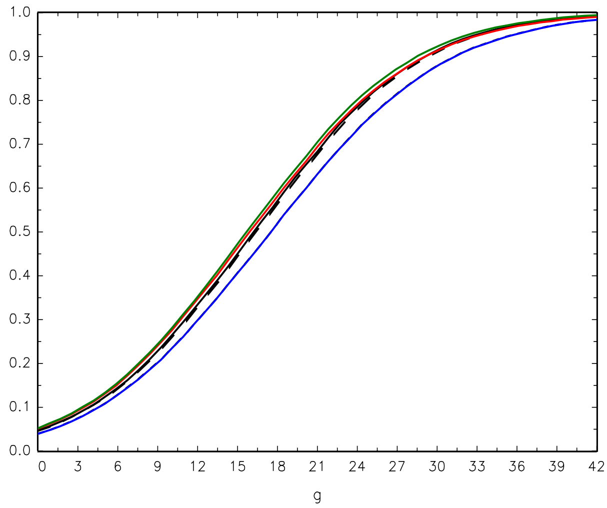
(a) $c = 0$



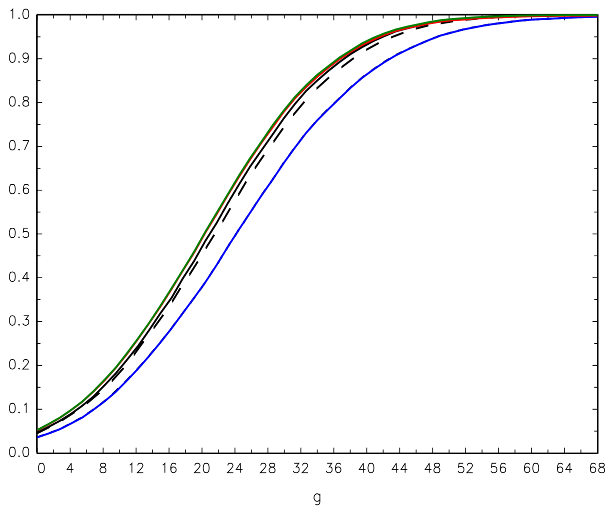
(b) $c = 10$



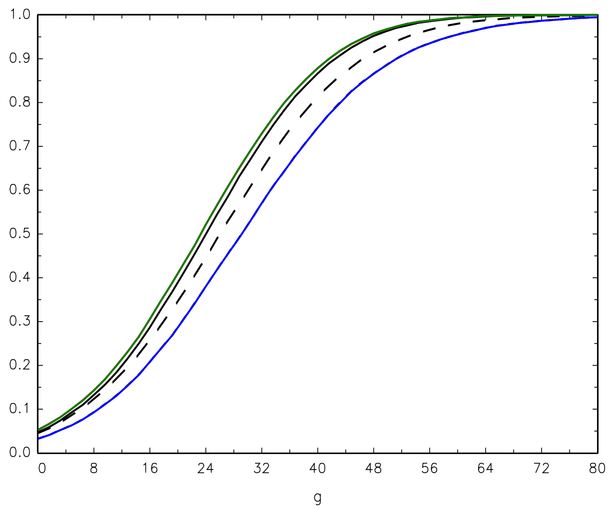
(c) $c = 25$



(d) $c = 50$

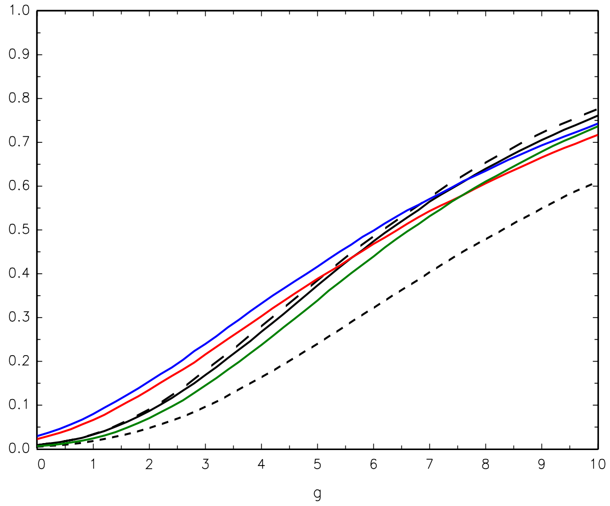


(e) $c = 100$

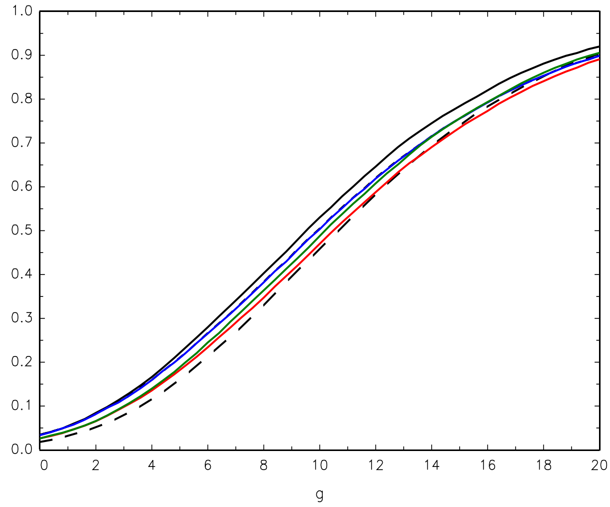


(f) $c = 200$

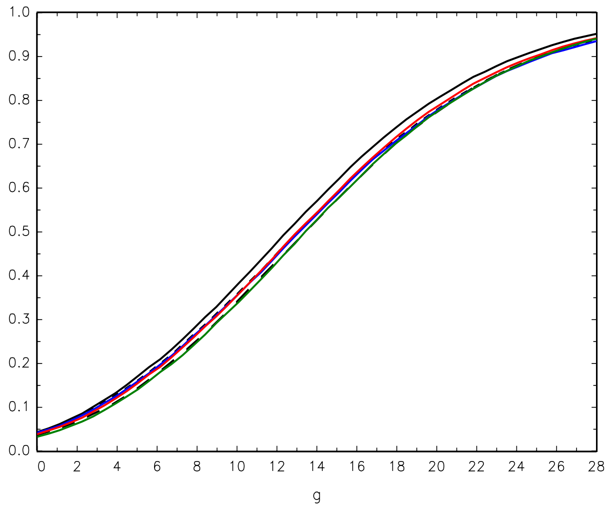
Figure 5. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0$; t_N: green solid line, t_C: black solid line, t'_C: short-dashed line, t^*_C: long-dashed line, t^w_C: blue solid line, t^{wt}_C: red solid line



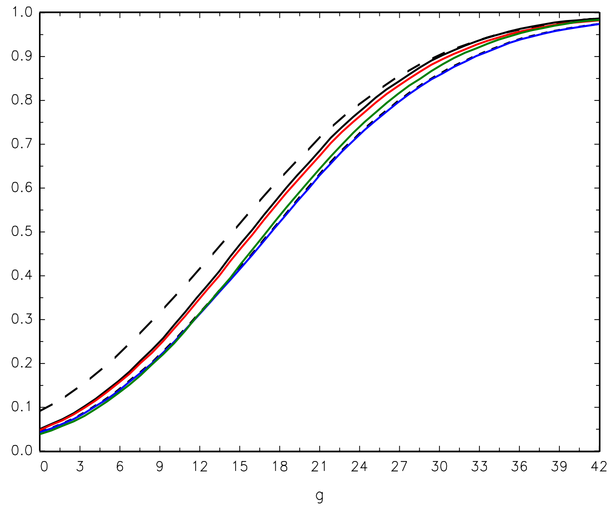
(a) $c = 0$



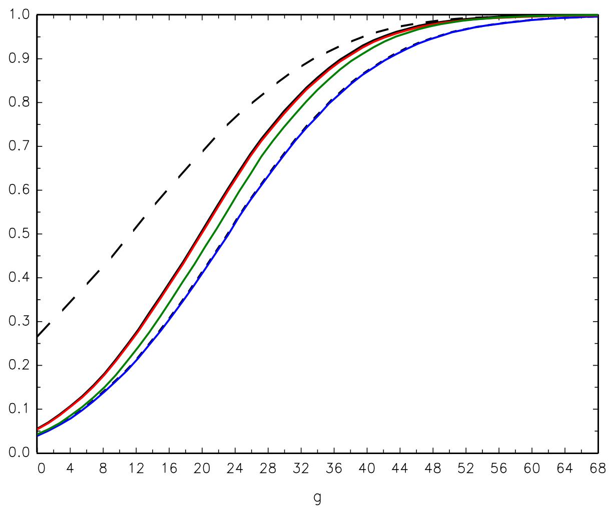
(b) $c = 10$



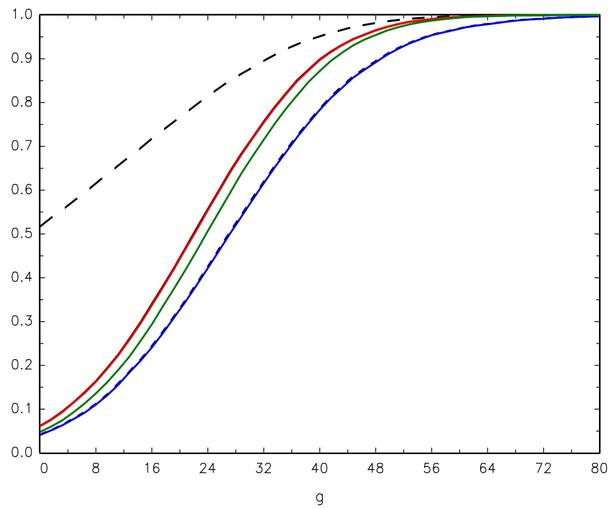
(c) $c = 25$



(d) $c = 50$

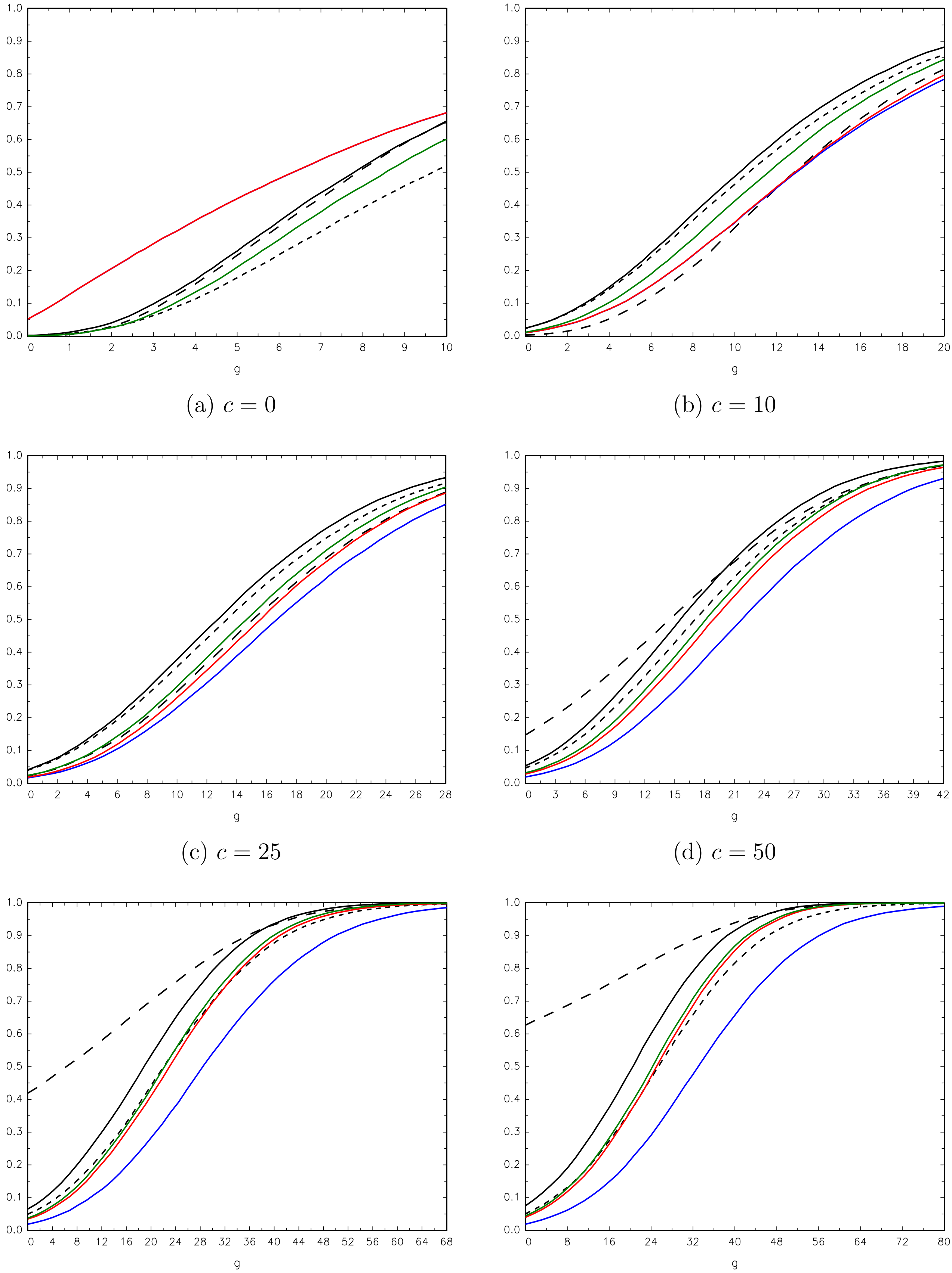


(e) $c = 100$



(f) $c = 200$

Figure 6. 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

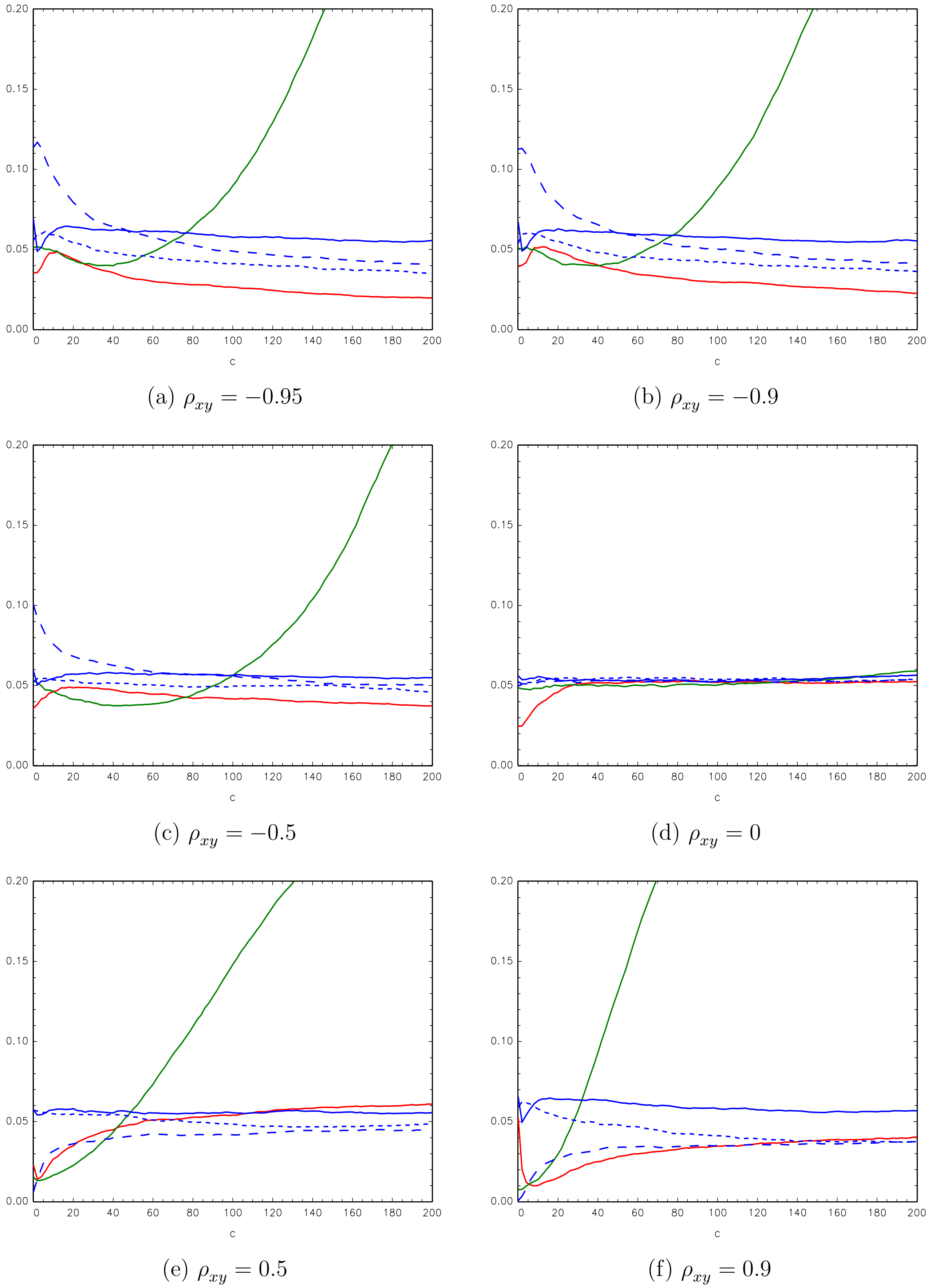


(a) $c = 0$ (b) $c = 10$

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

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

Figure 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 dashes, t^*_C: long dashes, t^w_C: blue solid, t^{wt}_C: red solid



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

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

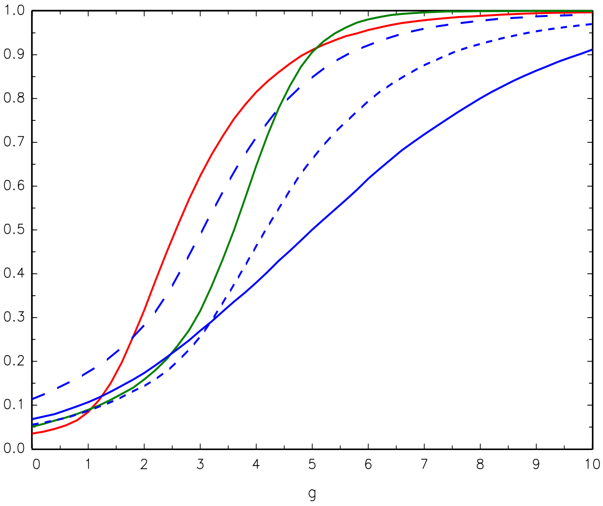
(c) $\rho_{xy} = -0.5$

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

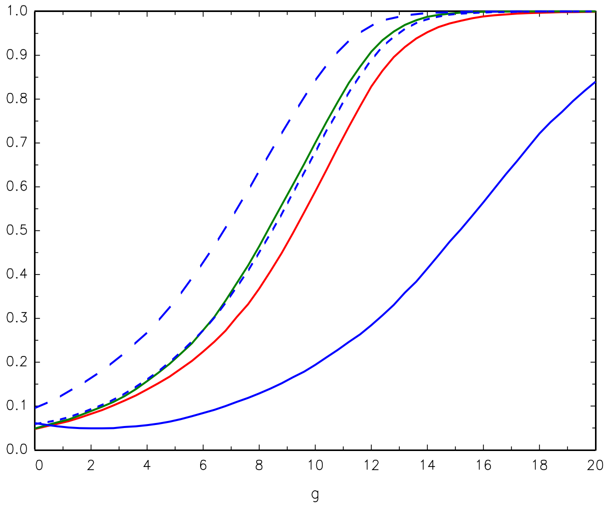
(e) $\rho_{xy} = 0.5$

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

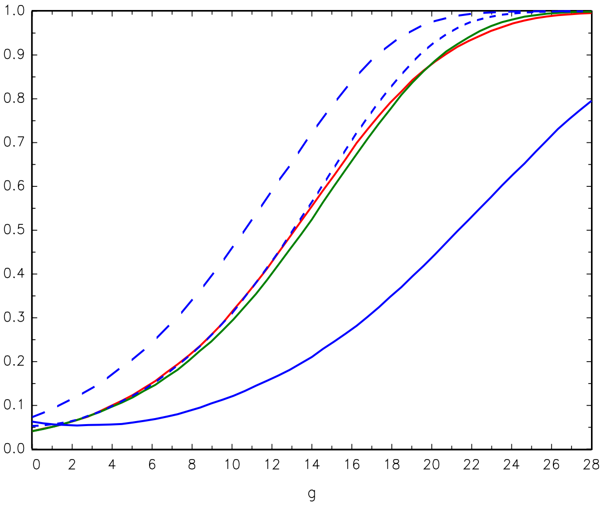
Figure 8. Finite sample size of nominal 0.05-level tests, $T = 200$;
t_C^{wt}: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line



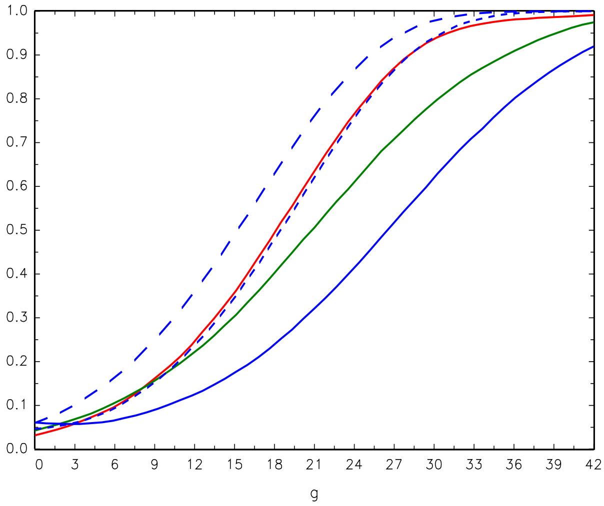
(a) $c = 0$



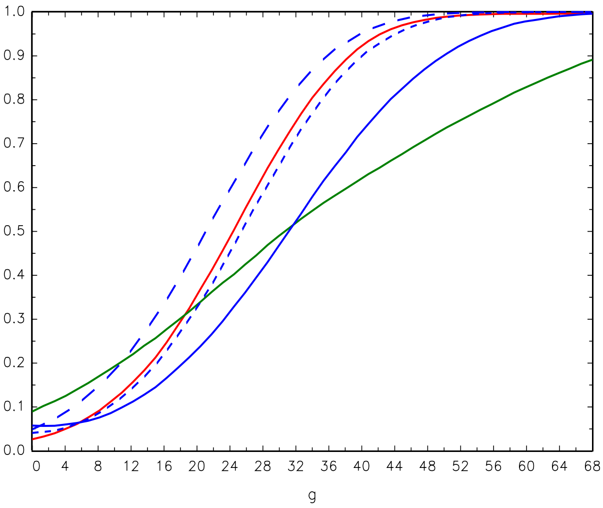
(b) $c = 10$



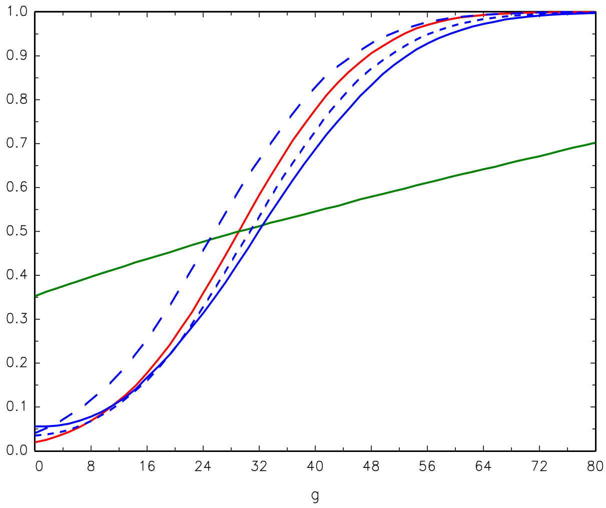
(c) $c = 25$



(d) $c = 50$



(e) $c = 100$



(f) $c = 200$

Figure 9. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.95$;
t_C^{wt}: —— , Q: —— , BD: —— , IVX_1: — — , IVX_2: - - -

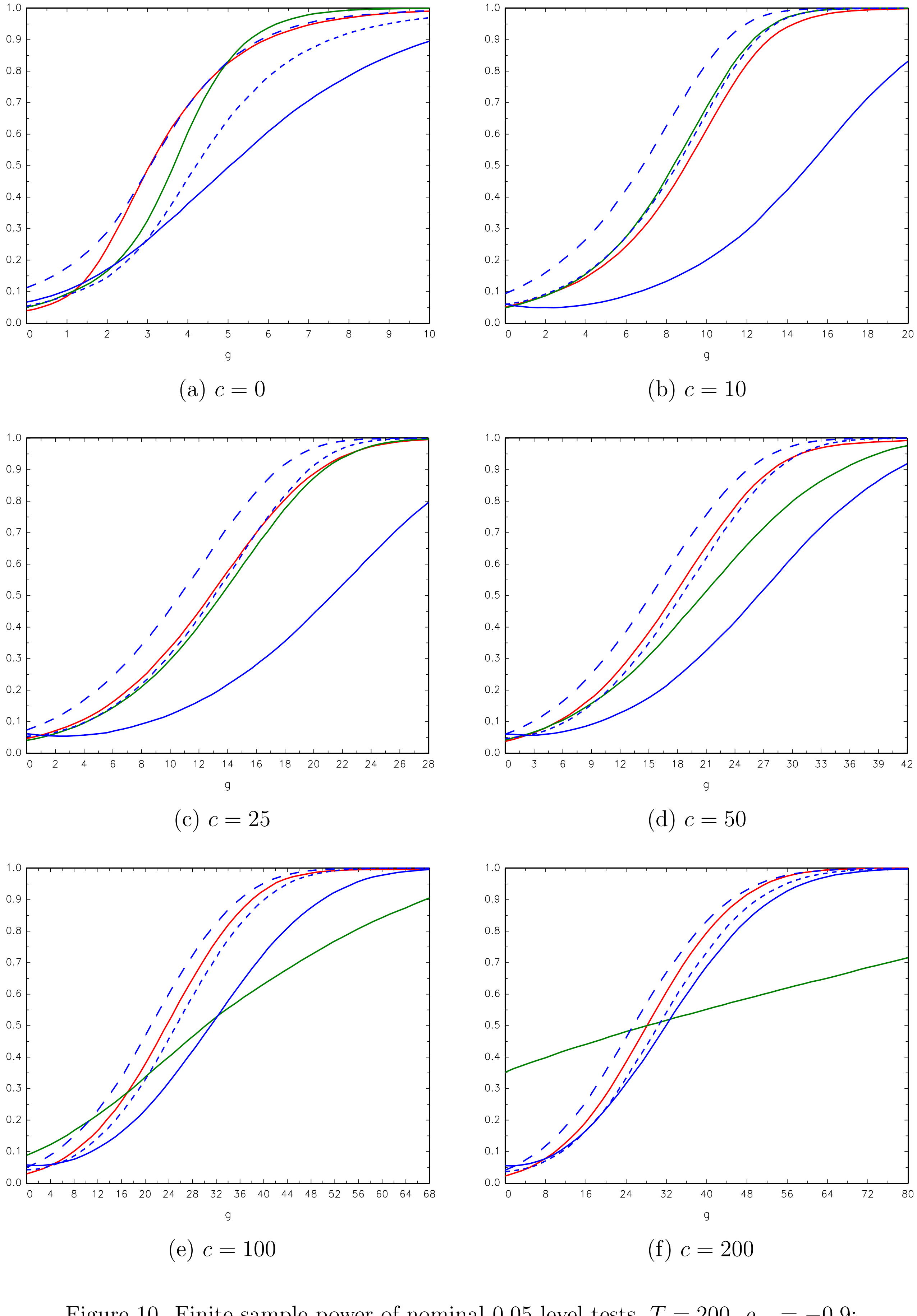
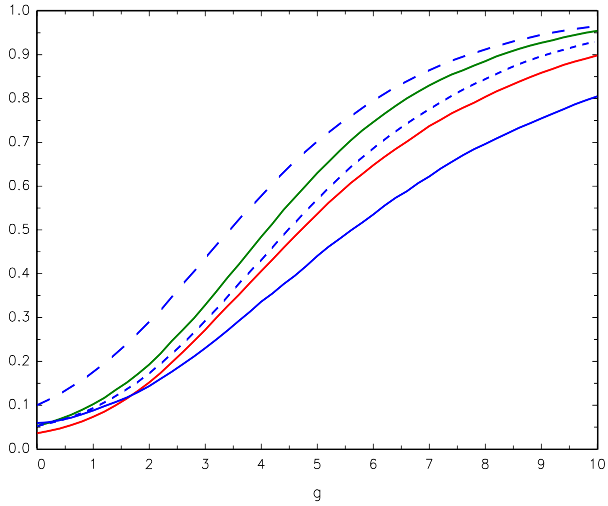
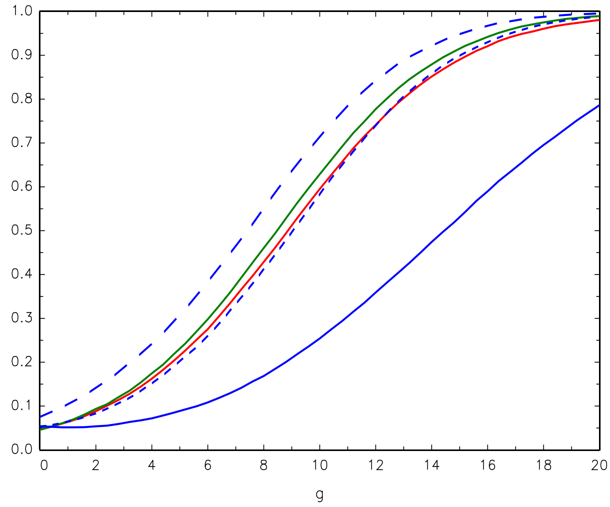


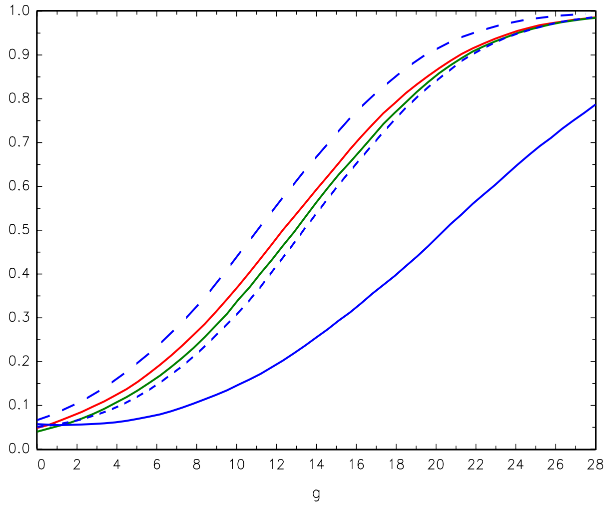
Figure 10. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = -0.9$;
t_C^{wt}: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line



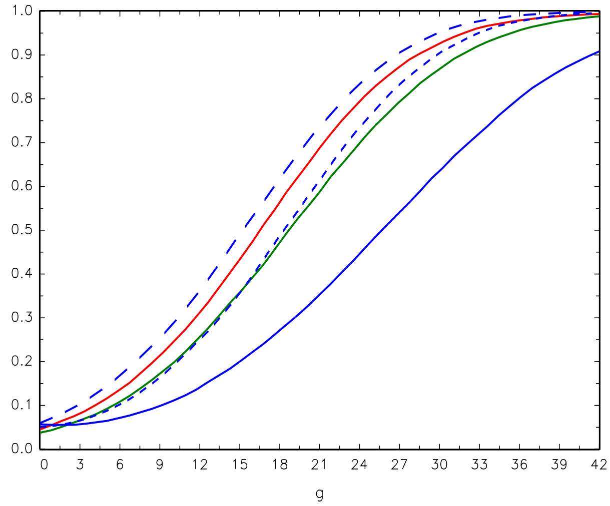
(a) $c = 0$



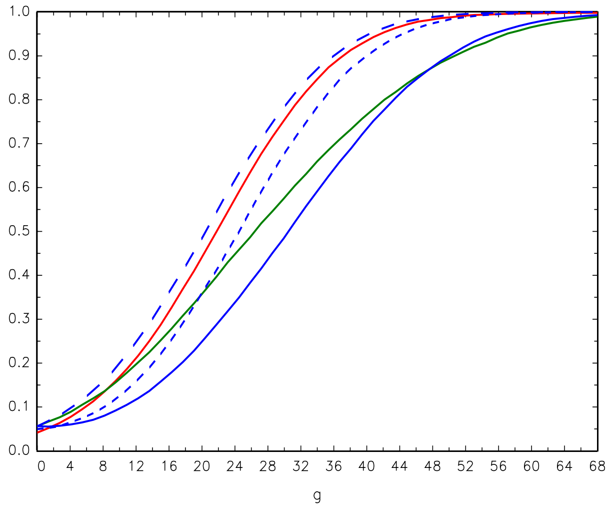
(b) $c = 10$



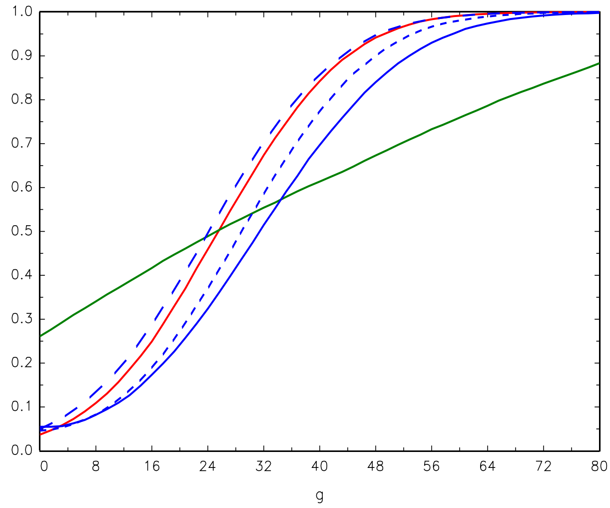
(c) $c = 25$



(d) $c = 50$

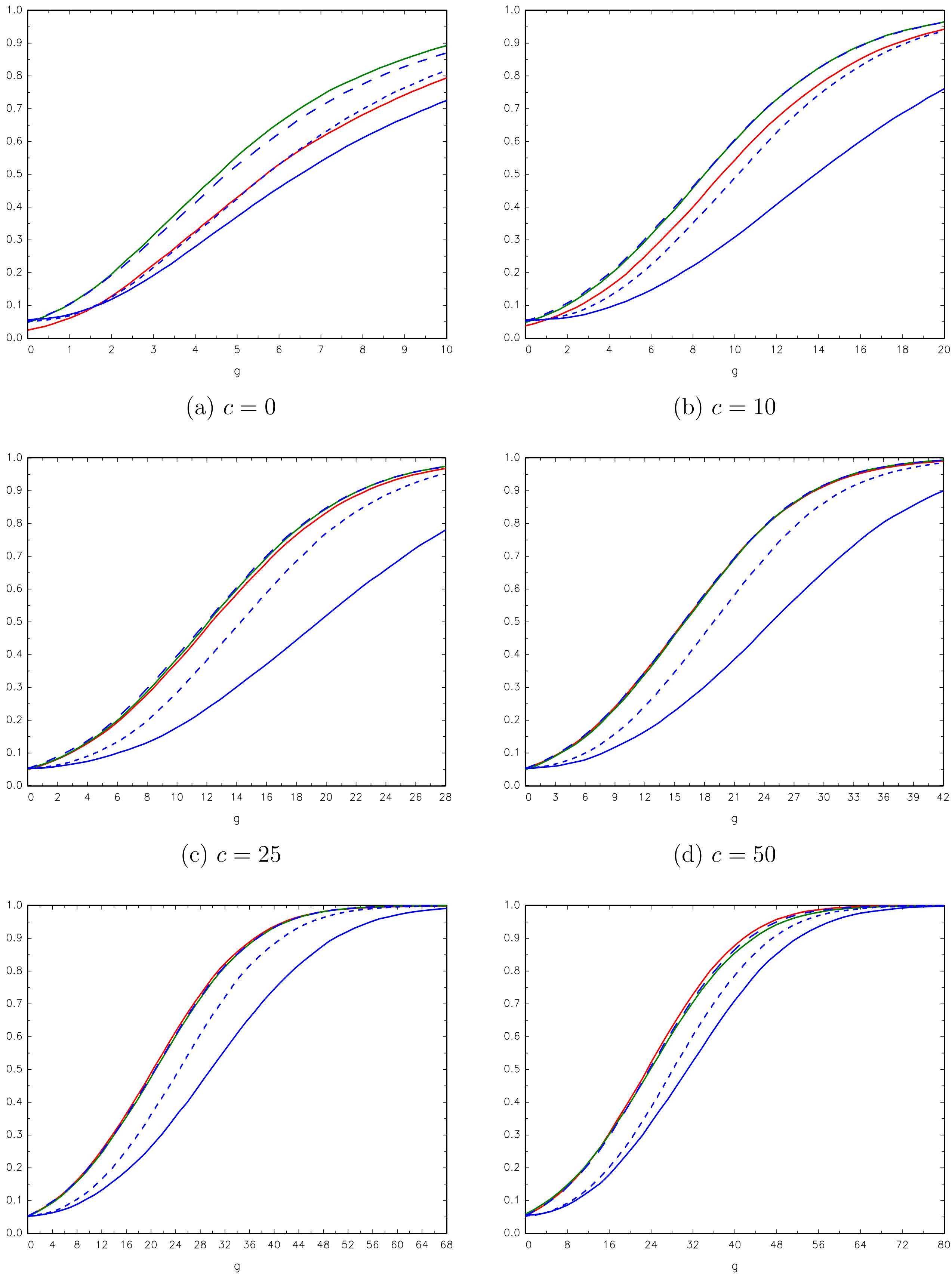


(e) $c = 100$



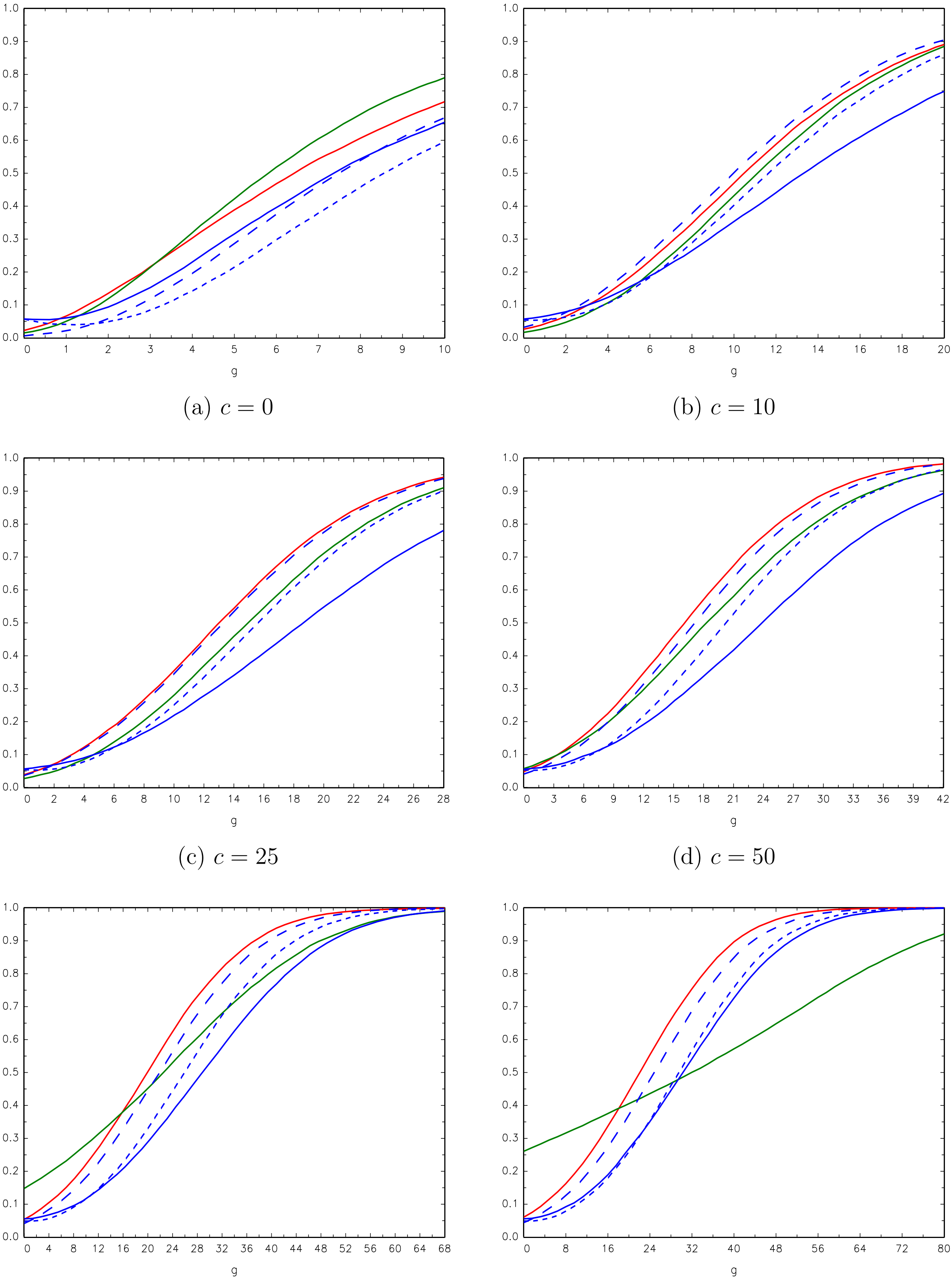
(f) $c = 200$

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}: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line

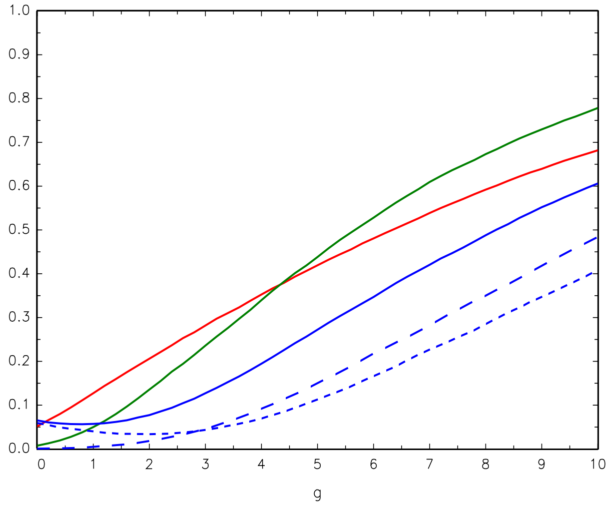


(a) $c = 0$ (b) $c = 10$

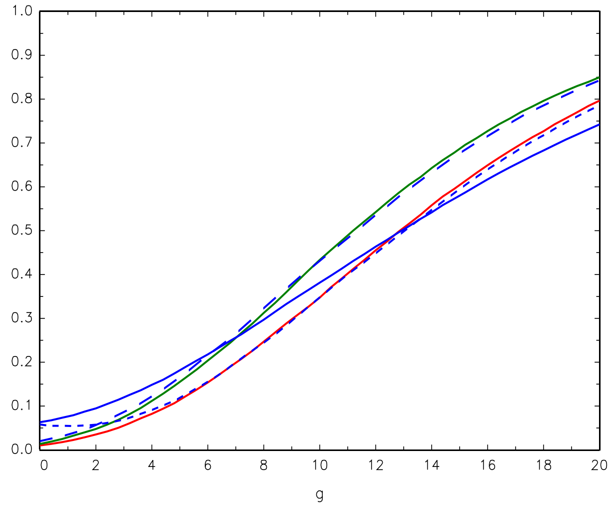
(c) $c = 25$ (d) $c = 50$

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

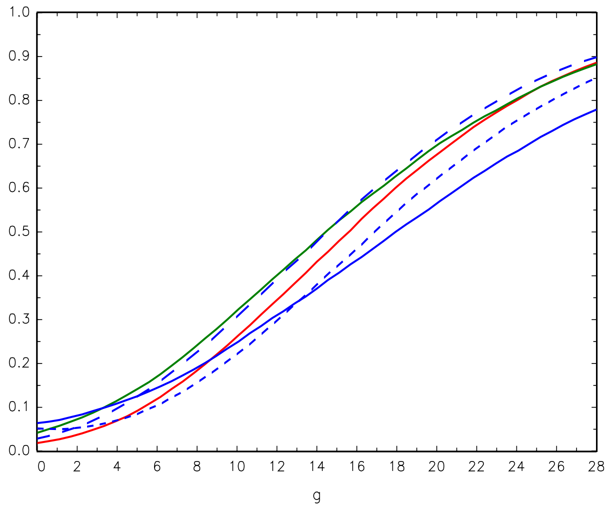
Figure 13. Finite sample power of nominal 0.05-level tests, $T = 200$, $\rho_{xy} = 0.5$; t_C^{wt}: red solid line, Q: green solid line, BD: blue solid line, IVX_1: blue long-dashed line, IVX_2: blue short-dashed line



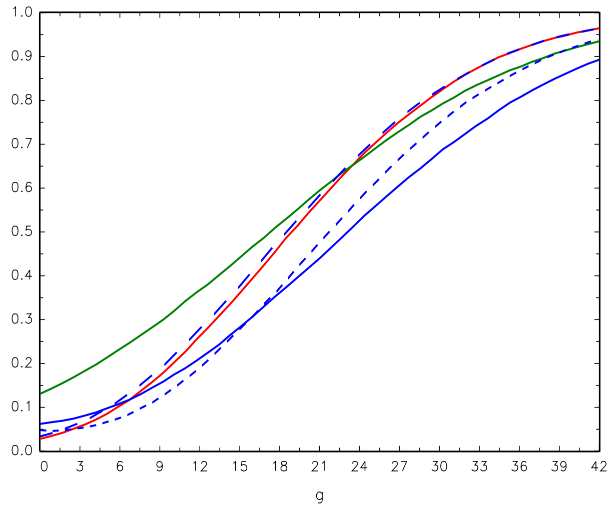
(a) $c = 0$



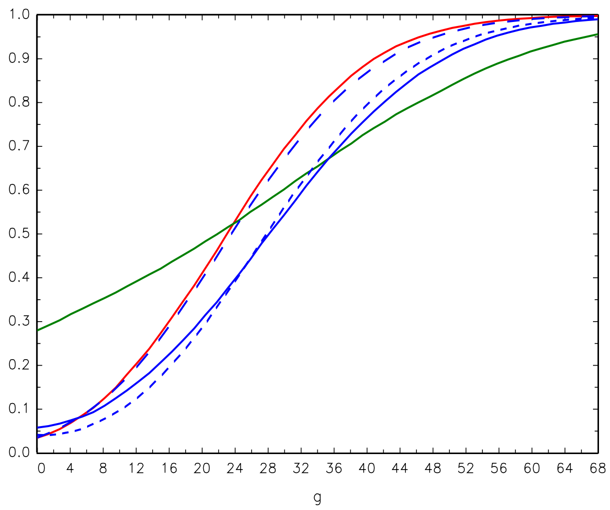
(b) $c = 10$



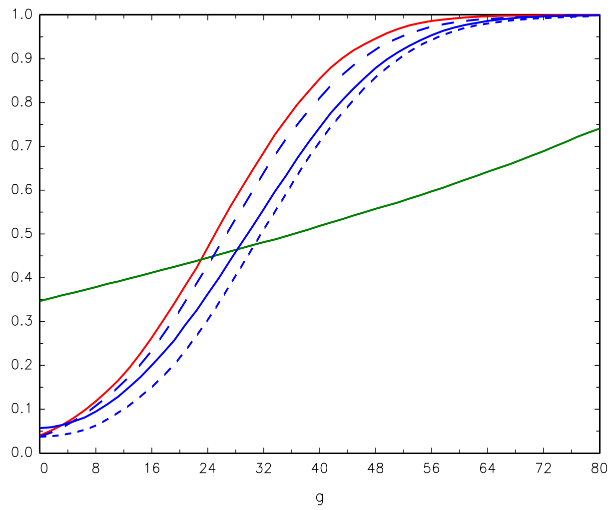
(c) $c = 25$



(d) $c = 50$



(e) $c = 100$



(f) $c = 200$

Figure 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: - - -