Discussion Papers
Department of Economics
University of Copenhagen

No. 17-09

Improved inference on cointegrating vectors in the presence of a near unit root using adjusted quantiles

Massimo Franchi
Søren Johansen

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

Improved inference on cointegrating vectors in the presence of a near unit root using adjusted quantiles*

Massimo Franchi†
Sapienza University of Rome

Søren Johansen‡
University of Copenhagen
and CREATES

April 21, 2017

Abstract

It is well known that inference on the cointegrating relations in a vector autoregression (CVAR) is difficult in the presence of a near unit root. The test for a given cointegration vector can have rejection probabilities under the null, which vary from the nominal size to more than 90%. This paper formulates a CVAR model allowing for many near unit roots and analyses the asymptotic properties of the Gaussian maximum likelihood estimator. Then a critical value adjustment suggested by McCloskey for the test on the cointegrating relations is implemented, and it is found by simulation that it eliminates size distortions and has reasonable power for moderate values of the near unit root parameter. The findings are illustrated with an analysis of a number of different bivariate DGPs.

Keywords: Long-run inference, test on cointegrating relations, likelihood inference, vector autoregressive model, near unit roots, Bonferroni type adjusted quantiles.

JEL Classification: C32.

1 Introduction

Elliot (1998) and Cavanagh, Elliott and Stock (1995) investigated the test on a coefficient of a cointegrating relation in the presence of a near unit root in a bivariate cointegrating regression. They show convincingly that when inference on the coefficient is performed as if the process has a unit root, then the size distortion is serious, see Figure 1 for a reproduction

*The first author gratefully acknowledges partial financial support from MIUR PRIN grant 2010J3LZEN. The second author is grateful to CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation, and to Adam McCloskey for discussions.

†Department of Statistics, Sapienza University of Rome. Email: massimo.franchi@uniroma1.it

‡Department of Economics, University of Copenhagen and CREATES, Department of Economics and Business, Aarhus University, DK-8000 Aarhus C. Address: Department of Economics University of Copenhagen, Øster Farimagsgade 5 Building 26, 1353 Copenhagen K, Denmark. E-mail: soren.johansen@econ.ku.dk.

of their results. This paper analyses the p-dimensional cointegrated VAR model with r cointegrating relations under local alternatives

$$\Delta y_t = (\alpha\beta' + T^{-1}\alpha_1 c\beta_1')y_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{1}$$

where α, β are $p \times r$ and ε_t is i.i.d. $N_p(0, \Omega)$. It is assumed that α_1 and β_1 are known $p \times (p-r)$ matrices of rank $p-r$, and c is $(p-r) \times (p-r)$ and an unknown parameter, such that the model allows for a whole matrix, c, of near unit roots. The matrix $\alpha\beta'$ describes a surface in the space of $p \times p$ matrices of dimension $p^2 - (p-r)^2$. The properties of the test Q_β can be very bad, when the actual data generating process (DGP) is a slight perturbation of the process generated by the model specified by $\alpha\beta'$. Thus, a model is formulated that in some particular "directions", given by the matrix $\alpha_1 c\beta_1'$, has a small perturbation of the order of T^{-1} and $(p-r)^2$ extra parameters, c, that are used to describe the near unit roots. A similar model could be suggested for near unit roots in the $I(2)$ model, see Di Iorio, Fachin and Lucchetti (2016), but this will not be attempted here.

The model defined by (1) contains as a special case the DGP analyzed by Elliot (1998). The likelihood ratio test, Q_β, for β equal to a given value, is derived assuming that $c = 0$ and analyzed when in fact near unit roots are present.

The parameters α, β, and Ω can be estimated consistently, but c cannot, and this is what causes the bad behaviour of Q_β.

The matrix $\alpha\beta' + T^{-1}\alpha_1 c\beta_1'$ has p^2 parameters, (α, β, c) so that the Gaussian maximum likelihood estimator in model (1) is least squares, and their limit distributions are found. The main contribution, however, is a simulation study for the bivariate VAR with $p = 2$, $r = 1$. It is shown that one of the methods introduced by McCloskey (2016, Theorem Bonf, p. 17) for allowing the critical value for Q_β to depend on the estimator of c, gives a much better solution to inference on β, in the case of a near unit root. The results of McCloskey (2016) also allow for multivariate parameters and for more complex adjustments, but in the present paper we focus for the simulations on the case with $p = 2$ and $r = 1$, so there is only one parameter in c.

The assumption that α_1 and β_1 are known is satisfied under the null, in the DGP analyzed by Elliot, see (25). This is of course convenient, because α_1, β_1 as free parameters, are not estimable.

Let θ denote the parameters α, β and Ω. For a given η (here 5% or 10%), the quantile $c_{\eta,\theta}(c)$ is defined by $P_{c,\theta}\{\hat{c} \leq c_{\eta,\theta}(c)\} = \eta$. Simulations show that the quantile is increasing in c, and solving the inequality for c, a $1-\eta$ confidence interval, $[0, c_{\eta,\theta}^{-1}(\hat{c})]$, is defined for c. For given ξ (here 90% or 95%) the quantile $q_{\xi,\theta}(c)$ is defined by $P_{c,\theta}\{Q_\beta \leq q_{\xi,\theta}(c)\} = \xi$, and McCloskey (2016) suggests replacing the critical value $q_{\xi,\theta}(c)$, by the stochastic critical value $q_{\xi,\theta}(c_{\eta,\theta}^{-1}(\hat{c}))$. This method is explained and implemented by a simulation study, and it is shown that it offers a solution to the problem of inference on β in the presence of a near unit root.

2 The vector autoregressive model with near unit roots

The model is given by (1) and the following standard $I(1)$ assumptions are made.

Assumption 1 *It is assumed that $r < p$, c is $(p-r) \times (p-r)$, and that the equation*

$$\det(I_p(1-z) - \alpha\beta' z) = 0$$

has $p-r$ roots equal to one, and the remaining roots are outside the unit circle, such that $|eigen(I_r+\beta'\alpha)|<1$. Moreover $\Pi=\alpha\beta'+T^{-1}\alpha_1 c\beta_1'$ has rank p and

$$\det(I_p(1-z)-\alpha\beta' z-T^{-1}\alpha_1 c\beta_1' z)=0$$

has all roots outside the unit circle.

In model (1) with cointegrating rank r and α_1 and β_1 known, the number of free parameters in α and β is $2pr-r^2=p^2-(p-r)^2$. Thus, allowing c to vary freely, there are $(p-r)^2$ extra parameters, and the maximum likelihood estimators for $\Pi=\alpha\beta'+T^{-1}\alpha_1 c\beta_1'$ and Ω are found by regression of Δy_t on y_{t-1}. The next theorem shows how the estimators for $\alpha,\beta,$ and c are calculated from $\hat{\Pi}$.

For any $p\times m$ matrix of rank $m<p$, we use the notation $a_\perp$ to indicate a $p\times(p-m)$ matrix of rank $p-m$, for which $a_\perp' a=0$, and the notation $\bar{a}=a(a'a)^{-1}$.

Theorem 1 *In model (1) with α_1 and β_1 known, the Gaussian maximum likelihood estimator of $\Pi=\alpha\beta'+T^{-1}\alpha_1 c\beta_1'$ is the coefficient in a least squares regression of Δy_t on y_{t-1}. For β normalized on $\beta'\beta_{1\perp}=I_r$, the maximum likelihood estimators are*

$$\hat{\alpha}=\hat{\Pi}\beta_{1\perp}, \tag{2}$$
$$\hat{\beta}'=(\alpha_{1\perp}'\hat{\Pi}\beta_{1\perp})^{-1}\alpha_{1\perp}'\hat{\Pi}, \tag{3}$$
$$\hat{c}=T(\beta_1'\hat{\Pi}^{-1}\alpha_1)^{-1}. \tag{4}$$

For $c=0$, such that the rank of Π is r, the test for a given value of β is

$$Q_\beta=T\log\frac{\det(S_{00}-S_{01}\beta(\beta' S_{11}^{-1}\beta)\beta' S_{10})}{\det(S_{00}-S_{01}\breve{\beta}(\breve{\beta}' S_{11}^{-1}\breve{\beta})\breve{\beta}' S_{10})}, \tag{5}$$

where the maximum likelihood estimator $\breve{\beta}$ is determined by reduced rank regression assuming the rank is r.

2.1 Asymptotic distributions

The basic asymptotic result for the analysis of the estimators and the test statistic is that $\alpha_\perp' y_t$ converges to an Ornstein-Uhlenbeck process. This technique was developed by Phillips (1988), and Johansen (1996, Chapter 14) is used as a reference for details related to the CVAR.

Under Assumption 1, the process given by (1) satisfies

$$T^{-1/2}\alpha_\perp' y_{[Tu]}\xrightarrow{D}K(u),$$

where K is the Ornstein-Uhlenbeck process

$$K(u)=\alpha_\perp'\int_0^u \exp\{\alpha_1 c\beta_1' C(u-s)\}dW_\varepsilon(s),$$

$C=\beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'$ and W_ε is Brownian motion generated by the cumulated ε_t.

Theorem 2 *The test Q_β for a given value of β, derived assuming $c = 0$, see (5), satisfies*

$$Q_\beta \xrightarrow{D} \chi^2_{(p-r)r} + B, \tag{6}$$

where the stochastic noncentrality parameter

$$B = tr\{\beta_1 c'\zeta c\beta_1'\beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\left(\int_0^1 KK'du\right)(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'\}, \tag{7}$$

is independent of the χ^2 distribution and has expectation

$$E(B) = tr\{\beta_1 c'\zeta c\beta_1' C\left(\int_0^1 (1-v)\exp(v\tau C)\Omega\exp(vC'\tau')dv\right)C'\}. \tag{8}$$

Here $\zeta = \alpha_1'\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}\alpha_1$ and $\tau = \alpha_1 c\beta_1'$, so it follows that $E(B) = 0$ if and only if $\alpha_1'\Omega^{-1}\alpha = 0$, in which case $Q_\beta \xrightarrow{D} \chi^2_{(p-r)r}$.

Let β be normalized as $\beta'\beta_{1\perp} = I_r$. The asymptotic distribution of the estimators, $\hat{\alpha}$, $\hat{\beta}$, $\hat{c}$, see (2), (3) and (4), are given as

$$T^{1/2}(\hat{\alpha} - \alpha) \xrightarrow{\mathrm{D}} N_{p\times r}(0, \Sigma_{\beta\beta}^{-1} \otimes \Omega), \tag{9}$$

$$T\beta_\perp'(\hat{\beta} - \beta) \xrightarrow{\mathrm{D}} \beta_\perp'\alpha_\perp\left(\int_0^1 KK'du\right)^{-1}\int_0^1 K(dW_\varepsilon)'\alpha_{1\perp}(\alpha'\alpha_{1\perp})^{-1}, \tag{10}$$

$$\hat{c} - c \xrightarrow{\mathrm{D}} (\alpha_\perp'\alpha_1)^{-1}\alpha_\perp'\int_0^1 (dW_\varepsilon)K'\left(\int_0^1 KK'du\right)^{-1}\alpha_\perp'\beta_\perp. \tag{11}$$

Note that the asymptotic distributions of $\hat{\beta}$ and $\hat{c}$ given in (10) and (11) are not mixed Gaussian, because $\alpha_{1\perp}'W_\varepsilon(u)$ is not independent of $K(u)$, which is generated by $\alpha_\perp'\varepsilon_t$. Note also that both $T\beta_\perp'(\hat{\beta} - \beta)$ and $\hat{c}$ have limit distributions that depend on the Dickey-Fuller type distribution $(\int_0^1 KK'du)^{-1}\int_0^1 K(dW_\varepsilon)'$.

Corollary 1 *In the special case where $r = p - 1$, we choose α_1 so that $c \geq 0$, and find*

$$E(B) = \frac{e^{2\delta c} - 1 - 2\delta c}{(2\delta)^2}\kappa\zeta, \tag{12}$$

where

$$\delta = \beta_1' C\alpha_1, \qquad \kappa = \beta_1' C\Omega C'\beta_1, \qquad \zeta = \alpha_1'\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}\alpha_1.$$

3 Critical value adjustment for test on β in the CVAR with near unit roots

In this section the method of McCloskey (2016) is illustrated by a number of simulation experiments. The simulations are performed with data generated by a bivariate model (1), where $p = 2$ and $r = 1$. The direction α_1 is chosen such that $c \geq 0$. The test Q_β for a given value of β, is calculated assuming $c = 0$, see (5). The simulations of Elliot (1998), see section

(3.1), show that there may be serious size distortions of the test, depending on the value of c and ρ, if the test is based on the quantiles from the asymptotic $\chi^2(1)$ distribution.

The method of McCloskey (2016, Theorem Bonf) consists in this case of replacing the $\chi^2(1)$ critical value, with a stochastic critical value depending on $\hat{c}$, in order to control the rejection probability under the null hypothesis.

Let $\theta = (\alpha, \beta, \Omega)$ and let $P_{c,\theta}$ denote the probability measure corresponding to the parameters c, θ. The method consists of finding the η quantile of $\hat{c}$, see (4), as defined by

$$P_{c,\theta}(\hat{c} \leq c_{\theta,\eta}(c)) = \eta,$$

for $\eta = 5\%$ or 10%, say, and the ξ quantile $q_{\theta,\xi}(c)$ of Q_β as defined by

$$P_{c,\theta}(Q_\beta \leq q_{\theta,\xi}(c)) = \xi,$$

for $\xi = 90\%$ or 95%, say.

The suggestion of McCloskey (2016) in this case is to construct by simulation, for a given θ and a grid of given of values $c \in (c_1, \ldots, c_n)$, the quantiles $c_{\theta,\eta}(c_i)$ and $q_{\theta,\xi}(c_i)$. It turns out, that both $c_{\theta,\eta}(c)$ and $q_{\theta,\xi}(c)$ are increasing in c, see Figure 3. Therefore, a solution $c_{\eta,\theta}^{-1}(\hat{c})$ can be found such that

$$P_{c,\theta}\{\hat{c} > c_{\eta,\theta}(c)\} = P_{c,\theta}\{c \leq c_{\eta,\theta}^{-1}(\hat{c})\} = 1 - \eta. \tag{13}$$

This gives a $1 - \eta$ confidence interval $[0, c_{\eta,\theta}^{-1}(\hat{c})]$ for c, based on the estimator $\hat{c}$. Note that for $c \leq c_{\eta,\theta}^{-1}(\hat{c})$ it holds that $q_{\theta,\xi}(c) \leq q_{\theta,\xi}(c_{\eta,\theta}^{-1}(\hat{c}))$, such that

$$1 - \xi = P_{c,\theta}\{Q_\beta > q_{\theta,\xi}(c)\} \geq P_{c,\theta}[Q_\beta > q_{\theta,\xi}\{c_{\eta,\theta}^{-1}(\hat{c})\}].$$

Hence, maximizing over $0 \leq c \leq c_{\eta,\theta}^{-1}(\hat{c})$, it is seen that

$$\max_{0 \leq c \leq c_{\eta,\theta}^{-1}(\hat{c})} P_{c,\theta}[Q_\beta > q_{\theta,\xi}\{c_{\eta,\theta}^{-1}(\hat{c})\}] \leq 1 - \xi.$$

McCloskey (2016) proved that under suitable conditions

$$1 - \xi - \eta \leq \limsup_{T \to \infty} \sup_{0 \leq c \leq c_{\eta,\theta}^{-1}(\hat{c})} P_{c,\theta}[Q_\beta > q_{\theta,\xi}\{c_{\eta,\theta}^{-1}(\hat{c})\}] \leq 1 - \xi.$$

Thus, the limiting rejection probability, for given θ, of the test on β, calculated as if $c = 0$, but replacing the $\chi_\xi^2(1)$ quantile by the stochastic quantile $q_{\theta,\xi}(c_{\eta,\theta}^{-1}(\hat{c}))$, lies between $1 - \xi - \eta$ and $1 - \xi$.

Note that for the theoretical analysis of the method, the parameters (θ, c, Ω), are assumed known in order to simulate the quantiles in the distribution of $\hat{c}$ and Q_β, for a range of c values, such that $c_{\eta,\theta}^{-1}(\hat{c})$ and the corrected quantile $q_{\theta,\xi}(c_{\eta,\theta}^{-1}(\hat{c}))$ can be found. It obviously simplify matters that in all the examples it turns out that $c_{\theta,\xi}(c)$ is approximately linear in c, and $q_{\theta,\xi}(c)$ is approximately quadratic in c, see Figure 3.

To implement the result in practice, however, one has of course to replace θ, Ω by a consistent estimator in the simulations.

3.1 The simulation study of Elliot (1998)

The DGP is defined by the equations,

$$y_{1t} = \left(1 - \frac{c}{T}\right) y_{1t-1} + u_{1t}, \tag{14}$$

$$y_{2t} = \gamma y_{1t} + u_{2t}. \tag{15}$$

It is assumed that $u_t = (u_{1t}, u_{2t})'$ are i.i.d. $N_2(0, \Omega_u)$ with

$$\Omega_u = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

and the initial values are $y_{10} = y_{20} = 0$. The data $y_1, \ldots, y_T$ are generated from (14) and (15), and the test statistic Q_β for the hypothesis $\gamma = 0$, is calculated using (5).

The DGP defined by (14) and (15) is contained in model (1) for $p = 2$. Note that $y_{2t} = \gamma(1 - c/T)y_{1t-1} + \gamma u_{1t} + u_{2t}$ such that

$$\alpha = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \ \beta = \begin{pmatrix} \gamma \\ -1 \end{pmatrix}, \ \alpha_1 = \begin{pmatrix} -1 \\ -\gamma \end{pmatrix}, \ \beta_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{16}$$

where the sign on α_1 has been chosen such that $c \geq 0$. Finally $\varepsilon_{1t} = u_{1t}$ and $\varepsilon_{2t} = u_{2t} + \gamma u_{1t}$, and therefore

$$\Omega = \begin{pmatrix} 1 & \rho + \gamma \\ \rho + \gamma & 1 + \gamma^2 + 2\gamma\rho \end{pmatrix}.$$

For $c = 0$, the process $y_t = (y_{1t}, y_{2t})'$ is $I(1)$ and $\gamma y_{1t} - y_{2t}$ is stationary, and if c/T is close to zero, y_t has a near unit root.

Applying Corollary 1 to the DGP (14)-(15), the expectation of the test statistic Q_β is found to be

$$E(Q_\beta) = p - 1 + \frac{e^{-2c} - 1 + 2c}{4} \frac{\rho^2}{1 - \rho^2}, \tag{17}$$

which increases approximately linearly in c.

Based on $N = 1000$ simulations of errors $u_1, \ldots, u_T$, $T = 100$, the data $y_1, \ldots, y_T$, are constructed from the DGP for each combination of the parameters

$$(\gamma, c, \rho) \in [-0.5 : (0.01) : 0.5] \times [1 : (1) : 20] \times [-0.9 : (0.1) : 0.9],$$

where $[a : (b) : c]$ indicates the interval from a to c with step b. Based on each simulation, $\hat{c}$ and the test Q_β for $\gamma = 0$ are calculated.

Figure 1 shows the rejection probabilities of the test Q_β as a function of (c, ρ), using the asymptotic critical value, $\chi^2_{0.95}(1) = 3.84$, for a nominal rejection probability of 5%. The rejection probability increases with $|\rho|$ and with c. When $c = 10$ (corresponding to an autoregressive coefficient of $c/T = 0.9$) and $|\rho| = 0.7$, the size of the test Q_β is around 50%, as found in Elliott (1998). The results are analogous across models with an unrestricted constant term, or with a constant restricted to the cointegrating space. In the paper by Elliot (1998) a number of tests are analyzed, and it was found that they were quite similar in their performance and similar to the above likelihood ratio test Q_β from the CVAR with rank equal to 1.

3.2 Redoing the simulations with adjusted quantiles for Q_β

Data are simulated as above and first the rank test statistic, Q_r, see Johansen (1996, Chapter 11) for rank equal to 1, is calculated. The rejection probabilities for Q_r are given in Figure 2 and they show that for $c = 20$, the hypothesis that the rank is 1, is practically certain to be rejected. If $c = 8$, the probability of rejecting that the rank is 1 is around 50%, so that plotting the rejection probabilities for $0 \leq c \leq 10$, covers the relevant values, see Figure 4.

For $\eta = 5\%$ and 10%, the quantiles $c_\eta(c)$ of $\hat{c}$ are reported in Figure 3 as a function of c. The quantiles $c_\eta(c)$ are nearly linear in c, and they are approximated by

$$\tilde{c}_\eta(c) = a_\eta + b_\eta c,$$

where the coefficients (a_η, b_η) depend on η, which is used to construct the upper confidence limit in (13) as

$$\tilde{c}_\eta^{-1}(\hat{c}) = (\hat{c} - a_\eta)b_\eta^{-1}.$$

For $\xi = 90\%$ and 95%, the quantiles $q_{\rho,\xi}(c)$ of Q_β are reported in Figure 3 as function of c for four values of ρ. It is seen that for given ρ, the quantiles $q_{\rho,\xi}(c)$ are quadratic in c, for relevant values of c, and hence they can be approximated by

$$\tilde{q}_{\rho,\xi}(c) = f_{\rho,\xi} + g_{\rho,\xi}c + h_{\rho,\xi}c^2, \tag{18}$$

where the coefficients $(f_{\rho,\xi}, g_{\rho,\xi}, h_{\rho,\xi})$ depend on ρ and ξ. The modified critical value is then constructed replacing (c, ρ) by $(\tilde{c}_\eta^{-1}(\hat{c}), \hat{\rho})$ in (18), and thus one finds the adjusted critical value

$$\tilde{q}_{\hat{\rho},\xi,\eta}(\hat{c}) = f_{\hat{\rho},\xi} + g_{\hat{\rho},\xi}(\hat{c} - a_\eta)b_\eta^{-1} + h_{\rho,\xi}((\hat{c} - a_\eta)b_\eta^{-1})^2 \tag{19}$$

which depends on estimated values, $\hat{c}$ and $\hat{\rho}$, and on discretionary values, ξ and η.

For $(\xi, \eta) = (95\%, 5\%)$, the rejection frequency of Q_β, the test for $\gamma = 0$, is calculated using the adjusted critical value in (19) and reported as a function of c for four values of ρ in Figure 4 together with the unadjusted rejection probabilities. In all cases the rejection frequency is in a neighborhood of the nominal size of 5%; hence the procedure is able to eliminate size-distortions almost completely for $c \leq 5$. The power of the test is shown in Figure 5 for values of $|\gamma| \leq 1/2$. It is seen that the better rejection probabilities in Figure 4 are achieved together with a reasonable power, again for $c \leq 5$, where the probability of rejecting the hypothesis of $r = 1$ is around 30%. Notice that the test becomes slightly biased, that is, the power function is not flat around the null $\gamma = 0$.

3.3 A few examples of other DGP

Four other data generating processes are defined in Table 1, to investigate the role of different choices of α_1 and β_1 for the results on improving the rejection probabilities for test on β under the null and alternative. The DGPs all have $\alpha = -\beta = (-1, 1)'/2$. The vectors α_1 and β_1 are chosen to investigate different positions of the near unit root in the DGP.

The choice of DGP turns out to be important also for the test, Q_r, for $r = 1$. In fact the probability of rejecting $r = 1$ is around 50% for DGP 1 if $c = 4$, for DGP 2 if $c = 20$,whereas for DGP 3 and 4 the 50% value value is 8.

The rejection probabilities in Figure 6 are plotted for $c \leq 10$, to cover the most relevant values.

Four DGPs allowing for near unit roots, $\Omega = I_2$			
1:	$\begin{pmatrix} -\frac{1}{4}+c/T & \frac{1}{4}+c/T \\ \frac{1}{4}+c/T & -\frac{1}{4}+c/T \end{pmatrix}$	3:	$\begin{pmatrix} -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{4}+c/T & -\frac{1}{4}+c/T \end{pmatrix}$
2:	$\begin{pmatrix} -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4}+c/T \end{pmatrix}$	4:	$\begin{pmatrix} -\frac{1}{4}+c/T & \frac{1}{4}+c/T \\ \frac{1}{4} & -\frac{1}{4} \end{pmatrix}$

Table 1: Four different DGPs given by $\alpha = -\beta = (-1,1)'/2$ which are the basis for the simulations of rejection probabilities for the adjusted test for $\beta = (-1,1)'/2$. The positions of c/T give the different α_1 and β_1 .

The results are summarized in Figures 6 and 7. It is seen that the conclusions from the study of the DGP analyzed by Elliot seem to be valid also for other DGPs. For moderate values of c, the adjusted test has a rejection probability in a neighborhood of the nominal size 5%, and the power curve looks reasonable for $c \leq 5$, although the test is slightly biased, except for DGP 1. For this DGP, $\alpha_1 = \beta_1 = (1,1)'$, $\Omega = I_2$, such that $\alpha_1'\Omega^{-1}\alpha = 0$, which means that the asymptotic distribution of Q_β is $\chi^2(1)$, see Theorem 2, despite the near unit root. It is seen from Figure 6, there is only moderate distortion of the rejection probability in this case and in Figure 7, the power curve is symmetric around $\gamma = 0$, so the test is approximately unbiased.

4 Conclusion

It has been demonstrated that for the DGP analyzed by Elliot (1998), it is possible to apply the method of McCloskey (2016) to adjust the critical value in such a way that the rejection probabilities of the test for β, are reasonably close to the nominal values for relevant values of c, that is, for which $c \leq 5$. By simulating the power of the test for β, it is seen that for $c \leq 5$, the test has a reasonable power, even though the test is slightly biased.

Some other DGPs have been investigated and similar results found. What remains to be done, of course, is to investigate by simulations if the values of ξ, η can be tuned to give a specified rejection probability under the null taking into account the power. What also remains to be done, is to see if the assumption that α_1 and β_1 are known, is helpful in applications.

5 References

Cavanagh C., G. Elliott and J.H. Stock (1995) Inference in Models with Nearly Integrated Regressors, *Econometric Theory* 11, 1131-1147.

F. Di Iorio, F., S. Fachin and R. Lucchetti (2016) Can you do the wrong thing and still be right? Hypothesis testing in I(2) and near-I(2) cointegrated VARs, *Applied Economics,* 48:38, 3665-3678, DOI: 10.1080/00036846.2016.1142660.

Elliott, G., (1998) On the robustness of cointegration methods when regressors almost have unit roots, *Econometrica* 66, 149–158.

Johansen S., (1996) *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford University Press, Oxford.

McCloskey A. (2016) Bonferroni-Based Size-Correction for Nonstandard Testing Problems. Available at SSRN: https://ssrn.com/abstract=2171912 or http: // dx.doi.org/10.2139 /ssrn.2171912

Phillips P.C.B., (1988) Regression theory for near integrated time series. *Econometrica* 56, 1021-1044.

6 Appendix

Proof of Theorem 1. The product moments of Δy_t, y_{t-1}, and ε_t are denoted

$$S_{11} = T^{-1}\sum_{t=1}^{T} y_{t-1}y_{t-1}', \qquad S_{10} = T^{-1}\sum_{t=1}^{T} y_{t-1}\Delta y_{t-1}', \qquad S_{1\varepsilon} = T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t'.$$

The maximum likelihood estimator of Π is $\hat{\Pi} = S_{01}S_{11}^{-1}$, and $\hat{\Omega} = S_{00} - S_{01}S_{11}^{-1}S_{10}$. From

$$\hat{\Pi} = \hat{\alpha}\hat{\beta}' + T^{-1}\alpha_1\hat{c}\beta_1',$$

it follows, post-multiplying by $\beta_{1\perp}$ and using the normalization $\beta'\beta_{1\perp} = I_r$, that

$$\hat{\alpha} = \hat{\Pi}\beta_{1\perp},$$

which shows (2). Pre-multiplying instead by $\alpha_{1\perp}'$ it follows, inserting $\hat{\alpha} = \hat{\Pi}\beta_{1\perp}$ that

$$\alpha_{1\perp}'\hat{\Pi} = \alpha_{1\perp}'\hat{\alpha}\hat{\beta}' = \alpha_{1\perp}'\hat{\Pi}\beta_{1\perp}\hat{\beta}',$$

which proves (3). Inserting the estimates it is found that

$$\hat{\Pi} = \hat{\alpha}\hat{\beta}' + T^{-1}\alpha_1\hat{c}\beta_1' = \hat{\Pi}\beta_{1\perp}(\alpha_{1\perp}'\hat{\Pi}\beta_{1\perp})^{-1}\alpha_{1\perp}'\hat{\Pi} + T^{-1}\alpha_1\hat{c}\beta_1'. \tag{20}$$

Next $\hat{\Pi}$ is decomposed using

$$\hat{\Pi} = \hat{\Pi}\beta_{1\perp}(\alpha_{1\perp}'\hat{\Pi}\beta_{1\perp})^{-1}\alpha_{1\perp}'\hat{\Pi} + \alpha_1(\beta_1'\hat{\Pi}^{-1}\alpha_1)^{-1}\beta_1', \tag{21}$$

which is proved by premultiplying (21) by $\alpha_{1\perp}'$ and $\beta_1'\hat{\Pi}^{-1}$. Subtracting (20) and (21) and multiplying by $\bar{\alpha}_1'$ and $\bar{\beta}_1$, it is seen that

$$(\beta_1'\hat{\Pi}^{-1}\alpha_1)^{-1} = \hat{c}/T,$$

which proves (4). If $c = 0$, the maximum likelihood estimator $\breve{\beta}$ can be determined by reduced rank regression, see Johansen (1996, Chapter 6). ■

Proof of Theorem 2. *Proof of (6) and* (7): The limit results for the product moments are given first, using the normalization matrix $C_T = (\beta, T^{-1/2}\alpha_\perp)$,

$$C_T'S_{11}C_T = \begin{pmatrix} \beta'S_{11}\beta & T^{-1/2}\beta'S_{11}\alpha_\perp \\ T^{-1/2}\alpha_\perp'S_{11}\beta & T^{-1}\alpha_\perp'S_{11}\alpha_\perp \end{pmatrix} \overset{\mathrm{D}}{\rightarrow} \begin{pmatrix} \Sigma_{\beta\beta} & 0 \\ 0 & \int_0^1 KK'du \end{pmatrix}, \tag{22}$$

$$T^{1/2}C_T'S_{1\varepsilon} = \begin{pmatrix} T^{1/2}\beta'S_{1\varepsilon} \\ T^{-1}\alpha_\perp'S_{1\varepsilon} \end{pmatrix} \overset{\mathrm{D}}{\rightarrow} \begin{pmatrix} N_{r\times p}(0, \Omega\otimes\Sigma_{\beta\beta}) \\ \int_0^1 K(dW_\varepsilon)' \end{pmatrix}. \tag{23}$$

The test for a known value of β is given in (5). It is convenient for the derivation of the limit distribution of Q_β, to normalize $\breve{\beta}$ on the matrix $\alpha(\beta'\alpha)^{-1}$, such that $\breve{\beta}'\alpha(\beta'\alpha)^{-1} = I_r$, and define $\breve{\theta} = (\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'(\breve{\beta}-\beta)$. This gives the representation

$$\breve{\beta}-\beta = \alpha_\perp(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'(\breve{\beta}-\beta) + \beta(\alpha'\beta)^{-1}\alpha'(\breve{\beta}-\beta) = \alpha_\perp\breve{\theta}.$$

The proof can be found in Elliot (1998), and is just sketched here. The estimator for θ for known α, Ω and $c = 0$, is given by the equation

$$T\breve{\theta} = (\alpha_\perp' T^{-1}S_{11}\alpha_\perp)^{-1}(\alpha_\perp' S_{1\varepsilon} + \alpha_\perp' T^{-1}S_{11}\beta_1 c\alpha_1')\alpha_\Omega,$$

where $\alpha_\Omega = \Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}$. The limit distribution of $T\breve{\theta}$ follows from (22) and (23) as follows. Because $T^{-1}\alpha_\perp' S_{11}\beta \xrightarrow{\mathrm{P}} 0$ it follows that

$$\begin{aligned}\alpha_\perp' T^{-1}S_{11}\beta_1 c\alpha_1' &= \alpha_\perp' T^{-1}S_{11}(\alpha_\perp(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp' + \beta(\alpha'\beta)^{-1}\alpha')\beta_1 c\alpha_1' \\ &\xrightarrow{\mathrm{D}} \left(\int_0^1 KK'du\right)(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'\beta_1 c\alpha_1',\end{aligned}$$

and from $\alpha_\perp' S_{1\varepsilon} \xrightarrow{\mathrm{D}} \int_0^1 K(dW_\varepsilon)$, it is seen that

$$T\breve{\theta} \xrightarrow{\mathrm{D}} \left(\int_0^1 KK'du\right)^{-1}\left(\int_0^1 K(dW_\varepsilon) + \left(\int_0^1 KK'du\right)(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'\beta_1 c\alpha_1'\right)\alpha_\Omega = U,$$

say. Conditional on K, the distribution of U is Gaussian with variance $(\alpha'\Omega^{-1}\alpha)^{-1}\otimes(\int_0^1 KK'du)^{-1}$ and mean $(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'\beta_1 c\alpha_1'\alpha_\Omega$. The information about θ satisfies

$$-T^{-2}I_{\theta\theta} = tr\{\Omega^{-1}\alpha(d\theta)'\alpha_\perp' S_{11}\alpha_\perp(d\theta)\alpha'\} \xrightarrow{\mathrm{D}} tr\{\alpha'\Omega^{-1}\alpha(d\theta)'\int_0^1 KK'du(d\theta)\},$$

and inserting U for $(d\theta)$ determines the asymptotic distribution of Q_β. Conditional on K, this has a noncentral $\chi^2((p-r)r)$ distribution with noncentrality parameter

$$B = tr\{(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'\beta_1 c'\zeta c\beta_1'\beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\int_0^1 KK'du\},$$

where $\zeta = \alpha_1'\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}\alpha_1$, which proves (7). The marginal distribution is therefore a noncentral χ^2 distribution with a stochastic noncentrality parameter, which is independent of the χ^2 distribution, as shown by Elliot (1998).

Proof of (8): For $\tau = \alpha_1 c\beta_1'$ it is seen that

$$\begin{aligned}&Etr\{(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'\beta_1 c'\zeta c\beta_1'\beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\int_0^1 KK'du\} \\ &= Etr\{\beta_1 c'\zeta c\beta_1' C\int_0^1\left(\int_0^u \exp(\tau C(u-s))dW(s)\right)\left(\int_0^u dW(t)'\exp(C'\tau'(u-t))\right)duC'\} \\ &= tr\{\beta_1 c'\zeta c\beta_1' C\int_0^1\left(\int_0^u \exp(\tau C(u-s))\Omega\exp(C'\tau'(u-s))ds\right)duC'\} \\ &= tr\{\beta_1 c'\zeta c\beta_1' C\left(\int_0^1 (1-v)\exp(v\tau C)\Omega\exp(vC'\tau')dv\right)C'\},\end{aligned}$$

which proves (8). Note that this expression is zero if and only if $\zeta = 0$, or $\alpha_1'\Omega^{-1}\alpha = 0$, in which case the asymptotic distribution of Q_β is χ^2.

Proof of (9) and (10):
It follows that $\hat{\Pi} = S_{01}S_{11}^{-1}$ can be expressed as

$$\begin{aligned}\hat{\Pi} &= \alpha\beta' + T^{-1}\alpha_1 c\beta_1' + S_{\varepsilon 1}S_{11}^{-1} \qquad (24)\\ &= \alpha\beta' + T^{-1}\alpha_1 c\beta_1' + T^{-1/2}(T^{1/2}S_{\varepsilon 1}C_T)(C_T'S_{11}C_T)^{-1}\left(\beta, T^{-1/2}\alpha_\perp\right)'\\ &= \alpha\beta' + T^{-1}\alpha_1 c\beta_1' + T^{-1/2}M_{1T}\beta' + T^{-1}M_{2T}\alpha_\perp',\end{aligned}$$

where, using (22) and (23),

$$M_{1T} \xrightarrow{\mathrm{D}} M_1 = N_{p\times r}(0, \Sigma_{\beta\beta}^{-1}\otimes\Omega), \qquad (25)$$

$$M_{2T} \xrightarrow{\mathrm{D}} M_2 = \int_0^1 dW_\varepsilon K'\left(\int_0^1 KK'du\right)^{-1}. \qquad (26)$$

From $\hat{\alpha} = \hat{\Pi}\beta_{1\perp}$, and using the normalization $\beta'\beta_{1\perp} = I_r$ such that $\alpha = \Pi\beta_{1\perp}$, it is seen that

$$T^{1/2}(\hat{\alpha}-\alpha) = T^{1/2}(\hat{\Pi}-\Pi)\beta_{1\perp} = (M_{1T}\beta' + T^{-1/2}M_{2T}\alpha_\perp')\beta_{1\perp} \xrightarrow{\mathrm{D}} M_1\beta'\beta_{1\perp} = M_1,$$

which proves (9). From $\hat{\beta}' = (\alpha_{1\perp}'\hat{\Pi}\beta_{1\perp})^{-1}\alpha_{1\perp}'\hat{\Pi}$ follows that

$$\begin{aligned}T(\hat{\beta}-\beta)'\beta_\perp &= T(\alpha_{1\perp}'\hat{\Pi}\beta_{1\perp})^{-1}\alpha_{1\perp}'(\hat{\Pi}-\alpha\beta')\beta_\perp\\ &= (\alpha_{1\perp}'\hat{\Pi}\beta_{1\perp})^{-1}\alpha_{1\perp}'(T^{1/2}M_{1T}\beta' + M_{2T}\alpha_\perp')\beta_\perp\\ &\xrightarrow{\mathrm{D}} (\alpha_{1\perp}'\alpha\beta'\beta_{1\perp})^{-1}\alpha_{1\perp}'M_2\alpha_\perp'\beta_\perp = (\alpha_{1\perp}'\alpha)^{-1}\alpha_{1\perp}'M_2\alpha_\perp'\beta_\perp,\end{aligned}$$

where $T^{1/2}M_{1T}\beta'\beta_\perp = 0$, $\alpha_{1\perp}'\alpha_1 c\beta_1' = 0$ and $\beta'\beta_{1\perp} = I_r$. This proves (10).

Proof of (11): To analyse the limit of the estimator, define

$$A_T = (T^{-1/2}\bar{\alpha}, \alpha_\perp) \text{ and } B_T = (T^{-1/2}\bar{\beta}, \beta_\perp),$$

and write

$$\hat{c} = T(\beta_1'\hat{\Pi}^{-1}\alpha_1)^{-1} = (\beta_1'B_T(A_T'T\hat{\Pi}B_T)^{-1}A_T'\alpha_1)^{-1}.$$

The expansion (24), and the limits (25) and (26) are then applied to give the limit results

$$\begin{aligned}T^{-1/2}\bar{\alpha}'(T\hat{\Pi})T^{-1/2}\bar{\beta} &= I_r + O(T^{-1}) + O_P(T^{-1}),\\ T^{-1/2}\bar{\alpha}'(T\hat{\Pi})\beta_\perp &= 0 + O(T^{-1/2}) + O_P(T^{-1/2}),\\ \alpha_\perp'(T\hat{\Pi})\bar{\beta}T^{-1/2} &= 0 + O(T^{-1/2}) + M_{1T},\\ \alpha_\perp'(T\hat{\Pi})\beta_\perp &= 0 + \alpha_\perp'\alpha_1 c\beta_1'\beta_\perp + \alpha_\perp'M_{2T}\alpha_\perp'\beta_\perp.\end{aligned}$$

Thus

$$A_T'(T\hat{\Pi})B_T \xrightarrow{\mathrm{D}} \begin{pmatrix} I_r & 0\\ M_1 & \alpha_\perp'\alpha_1 c\beta_1'\beta_\perp + \alpha_\perp'M_2\alpha_\perp'\beta_\perp\end{pmatrix},$$

$$\begin{aligned}&B_T(A_T'(T\hat{\Pi})B_T)^{-1}A_T' \xrightarrow{\mathrm{D}} (0,\beta_\perp)\begin{pmatrix} I_r & 0\\ M_1 & \alpha_\perp'\alpha_1 c\beta_1'\beta_\perp + \alpha_\perp'M_2\alpha_\perp'\beta_\perp\end{pmatrix}^{-1}(0,\alpha_\perp)'\\ &= \beta_\perp(\alpha_\perp'\alpha_1 c\beta_1'\beta_\perp + \alpha_\perp'M_2\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'.\end{aligned}$$

Thus, multiplying by β_1' and α_1 and inverting, it is seen that because $\beta_1'\beta_\perp$ and $\alpha_1'\alpha_\perp$ are $(p-r)\times(p-r)$ of full rank,

$$\begin{aligned}\hat{c} = (\beta_1' B_T (A_T' T\hat{\Pi} B_T)^{-1} A_T'\alpha_1)^{-1} &\xrightarrow{\mathrm{D}} [\beta_1'\beta_\perp(\alpha_\perp'\alpha_1 c\beta_1'\beta_\perp + \alpha_\perp' M_2 \alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\alpha_1]^{-1} \\ &= (\alpha_\perp'\alpha_1)^{-1}(\alpha_\perp'\alpha_1 c\beta_1'\beta_\perp + \alpha_\perp' M_2\alpha_\perp'\beta_\perp)(\beta_1'\beta_\perp)^{-1} \\ &= c + (\alpha_\perp'\alpha_1)^{-1}\alpha_\perp' M_2\alpha_\perp'\beta_\perp(\beta_1'\beta_\perp)^{-1},\end{aligned}$$

which proves (11). ■

Proof of Corollary 1. *Proof of (12):* If $r = p-1$, the expression (8) can be reduced as follows. For $\tau = \alpha_1 c\beta_1'$

$$(\tau C)^2 = c\alpha_1 c(\beta_1' C\alpha_1)\beta_1' C = c(\beta_1' C\alpha_1)\alpha_1 c\beta_1' C = c\delta\tau C,$$

for $\delta = \beta_1' C\alpha_1$, and in general for $n \geq 0$, it is seen that

$$(\tau C)^{n+1} = (c\delta)^n \tau C.$$

Therefore, using $\beta_1 c'\zeta c\beta_1' = \beta_1 c'\tau\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}\tau'$,

$$E(B) = tr\{\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}\tau C\left(\int_0^1 (1-v)\exp(v\tau C)\Omega\exp(vC'\tau')dv\right)C'\tau'\}.$$

The integral can be calculated by the expansion

$$\begin{aligned}&\tau C\exp(v\tau C)\Omega\exp(vC'\tau')C'\tau' = \sum_{n,m=0}^{\infty}\frac{v^n}{n!}(\tau C)^{n+1}\Omega(C'\tau')^{m+1}\frac{v^m}{m!} \\ &= \sum_{n,m=0}^{\infty}\frac{(vc\delta)^{n+m}}{n!m!}\tau C\Omega C'\tau' = \exp(2vc\delta)c^2\kappa\alpha_1\alpha_1',\end{aligned}$$

where $\kappa = \beta_1' C\Omega C'\beta_1$. This allows the integral to be calculated

$$\begin{aligned}&\tau C\left(\int_0^1(1-v)\exp(v\tau C)\Omega\exp(vC'\tau')dv\right)C'\tau' \\ &= \left(\int_0^1(1-v)\exp(2vc\delta)dv\right)c^2\kappa\alpha_1\alpha_1' = \frac{e^{2\delta c}-1-2c\delta}{(2\delta c)^2}c^2\kappa\alpha_1\alpha_1'.\end{aligned}$$

Therefore

$$E(B) = \frac{e^{2c\delta}-1-2c\delta}{(2\delta)^2}\kappa\alpha_1'\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}\alpha_1 = (e^{2c\delta}-1-2c\delta)\frac{\kappa\zeta}{(2\delta)^2},$$

where $\zeta = \alpha_1'\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}\alpha_1$. ■

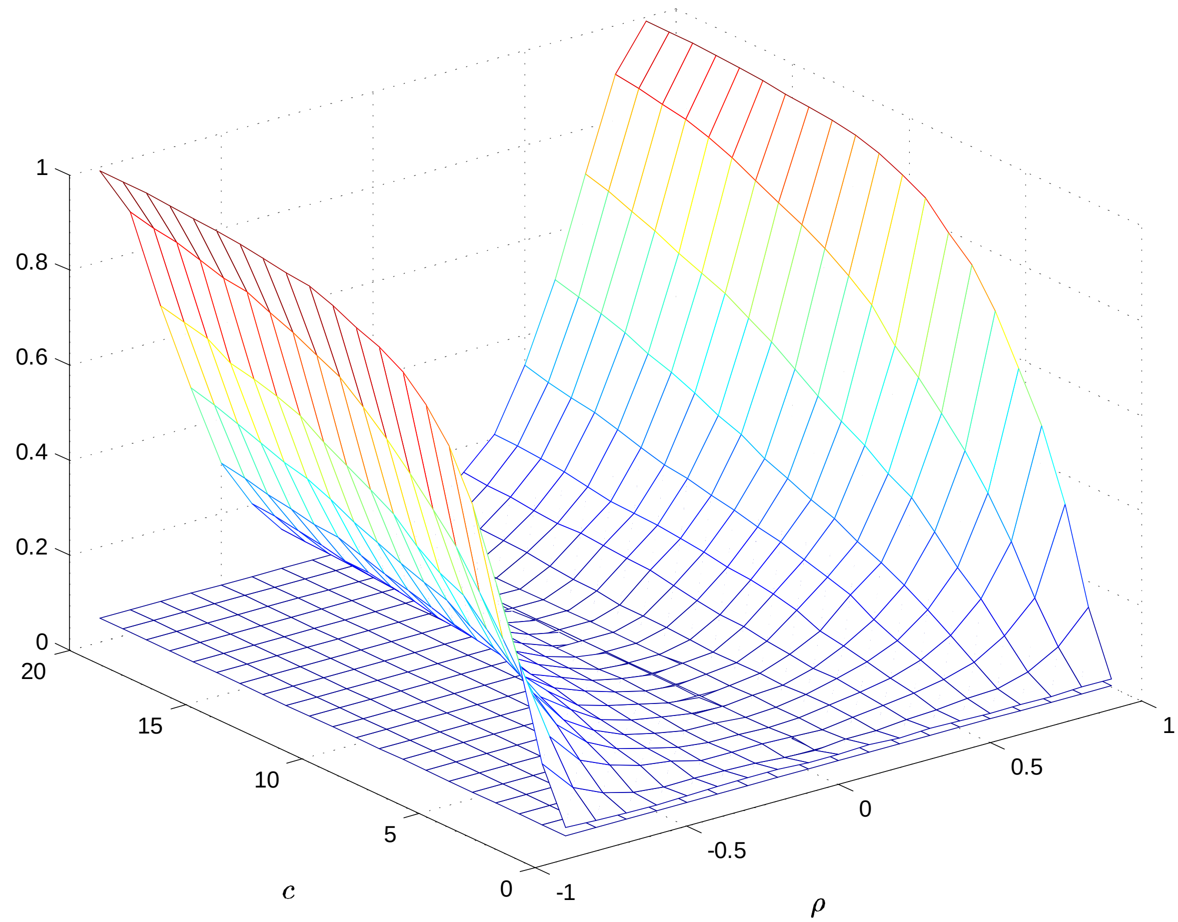

Figure 1: Rejection frequency of the test Q_β for $\gamma = 0$ using the $\chi^2(1)_{0.95}$ quantile as a function of c and ρ; $N = 1000$ simulations of $T = 100$ observations from the DGP (14) and (15).

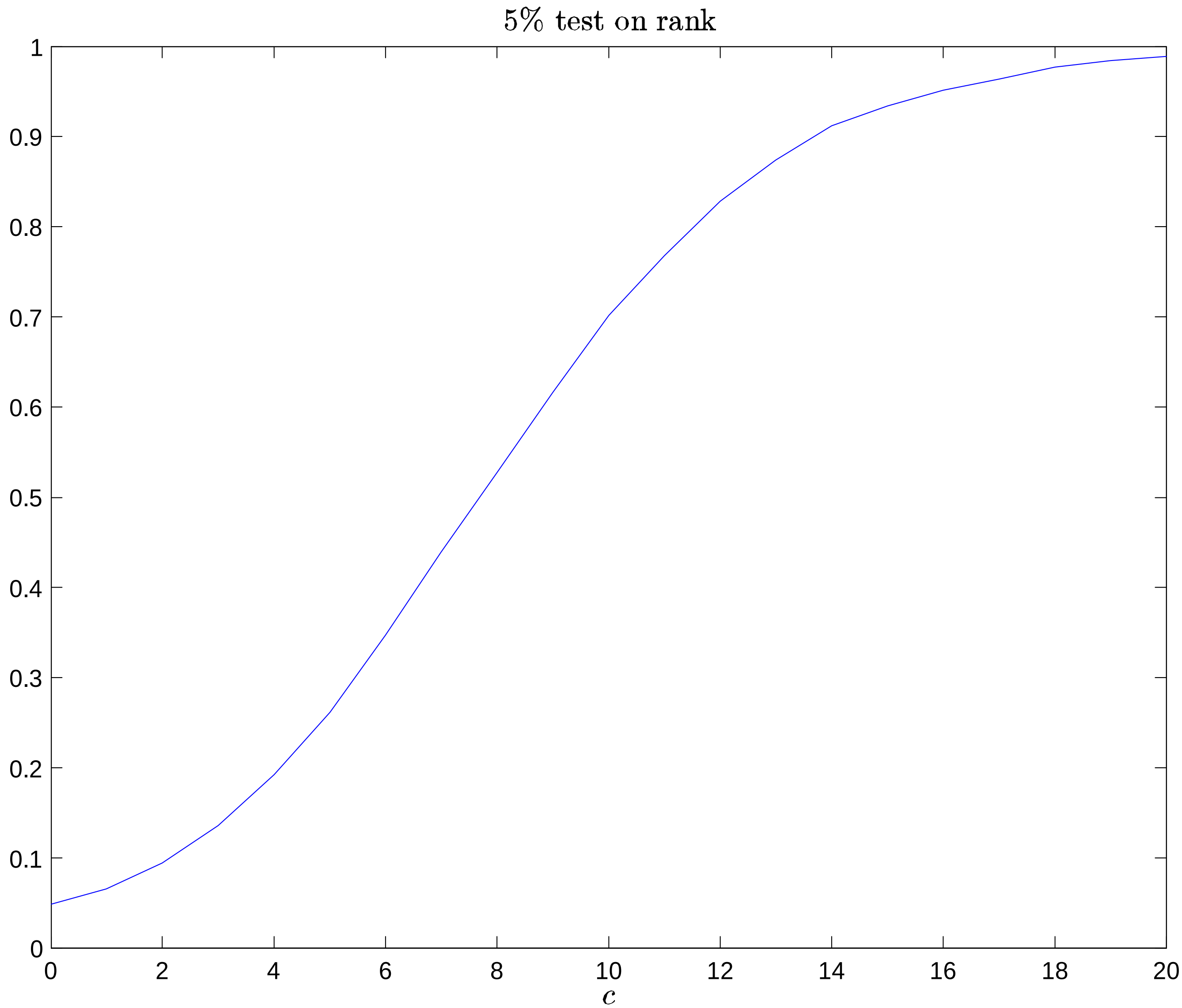

Figure 2: Rejection frequency of the test Q_r for $r = 1$ using Table 15.1 in Johansen (1996) as a function of c; $N = 1000$ simulations of $T = 100$ observations from the DGP (14) and (15).

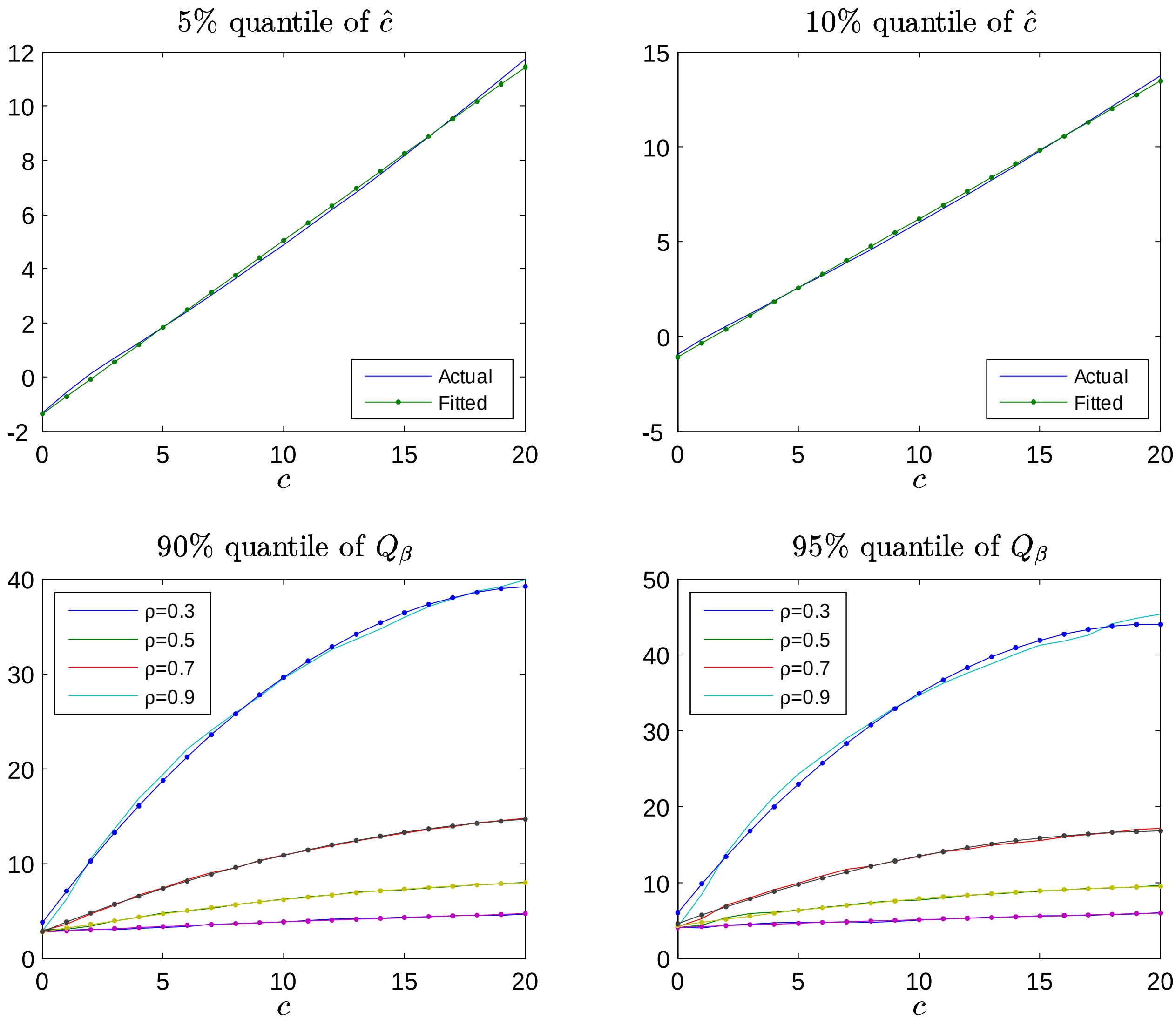

Figure 3: Quantiles and fitted values in the distributions of $\hat{c}$ and Q_β as a function of c for different values of ρ; $N = 1000$ simulations of $T = 100$ observations from the DGP (14) and (15).

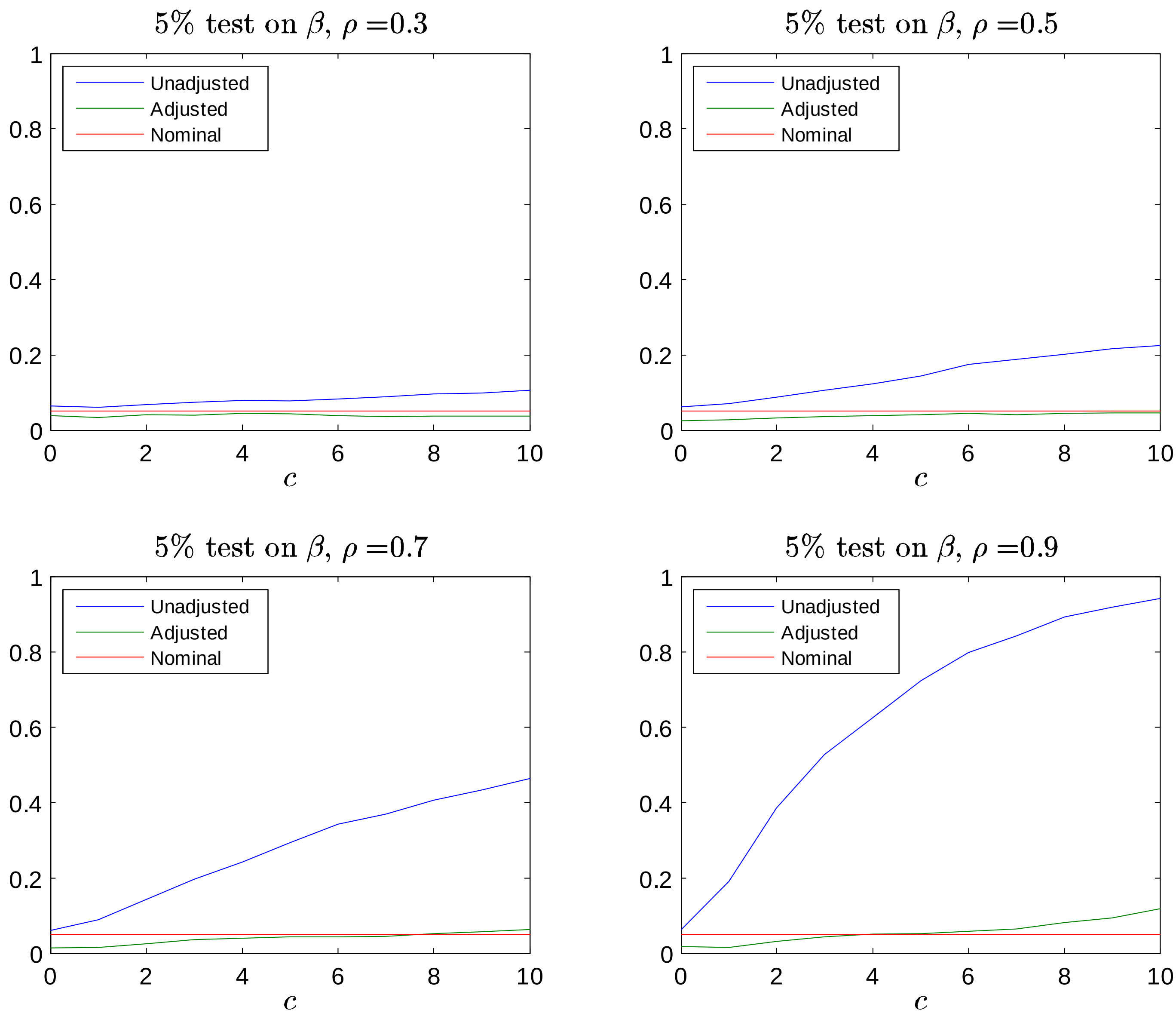

Figure 4: Rejection frequency of the test Q_β for $\gamma = 0$ using the $\chi^2(1)_{0.95}$ quantile (unadjusted) and the adjusted quantile in (19) for $\xi = 95\%$ and $\eta = 5\%$, as a function of c for different values of ρ; $N = 1000$ simulations of $T = 100$ observations from the DGP (14) and (15).

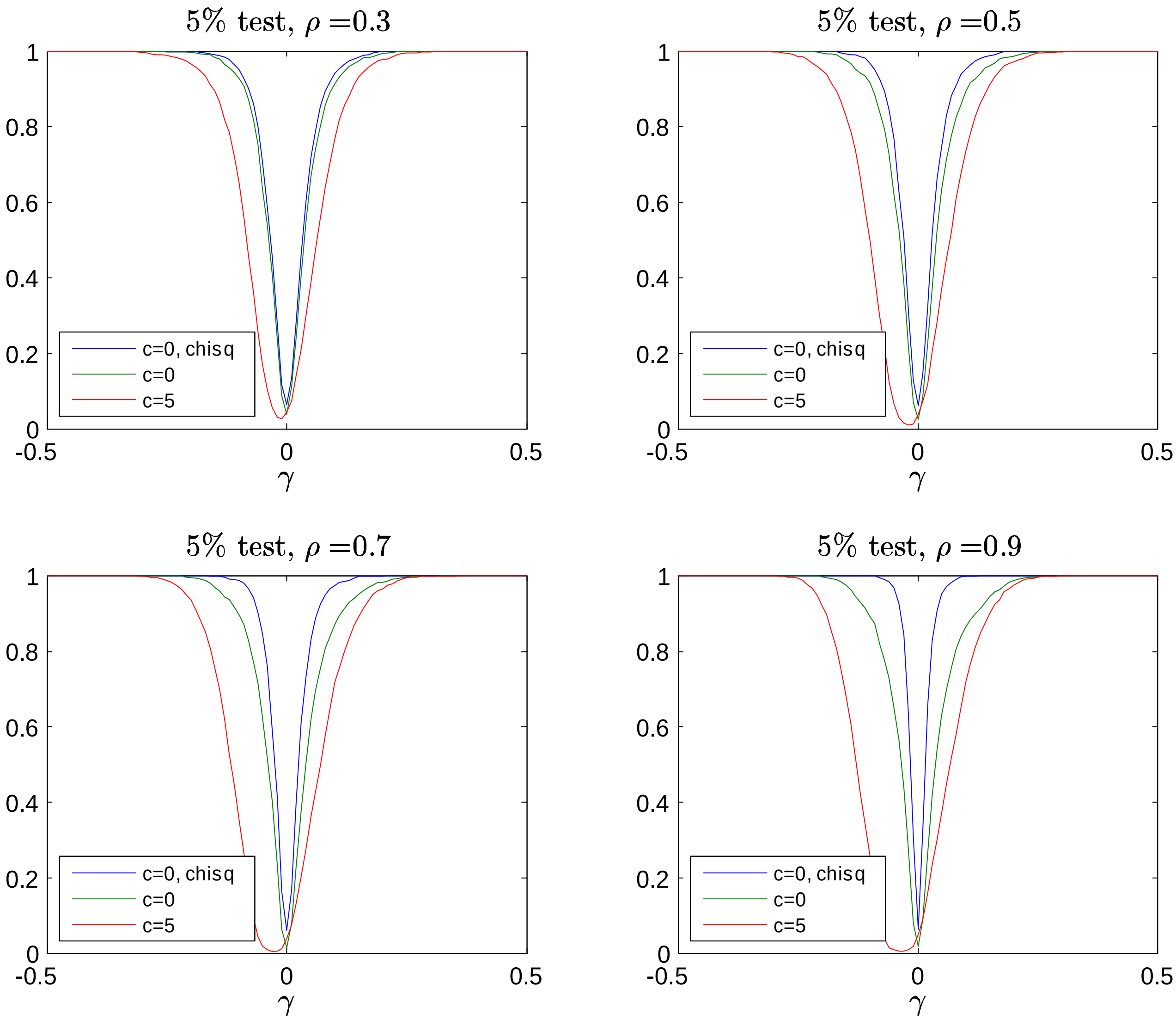

Figure 5: Rejection frequency of the test Q_β for $\gamma = 0$ using the $\chi^2(1)_{0.95}$ quantile (unadjusted) and the adjusted quantile in (19) for $\xi = 95\%$ and $\eta = 5\%$, as a function of γ for different values of c and ρ; $N = 1000$ simulations of $T = 100$ observations from the DGP (14) and (15).

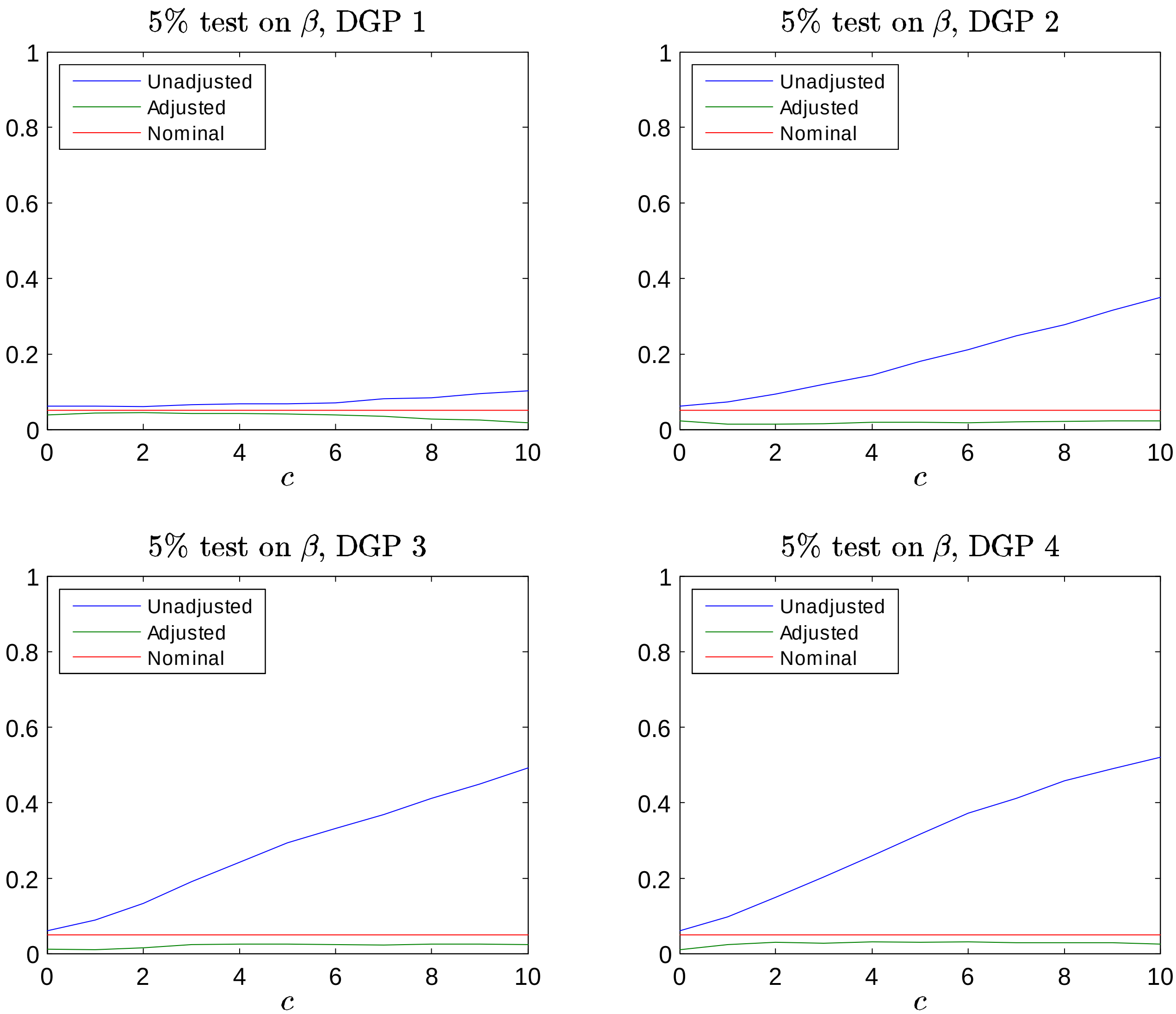

Figure 6: Rejection frequency of the test Q_β for $\gamma = 0$ using the $\chi^2(1)_{0.95}$ quantile (unadjusted) and the adjusted quantile in (19) $\xi = 95\%$ and $\eta = 5\%$, as a function of c; $N = 1000$ simulations of $T = 100$ observations from the DGPs in Table 1.

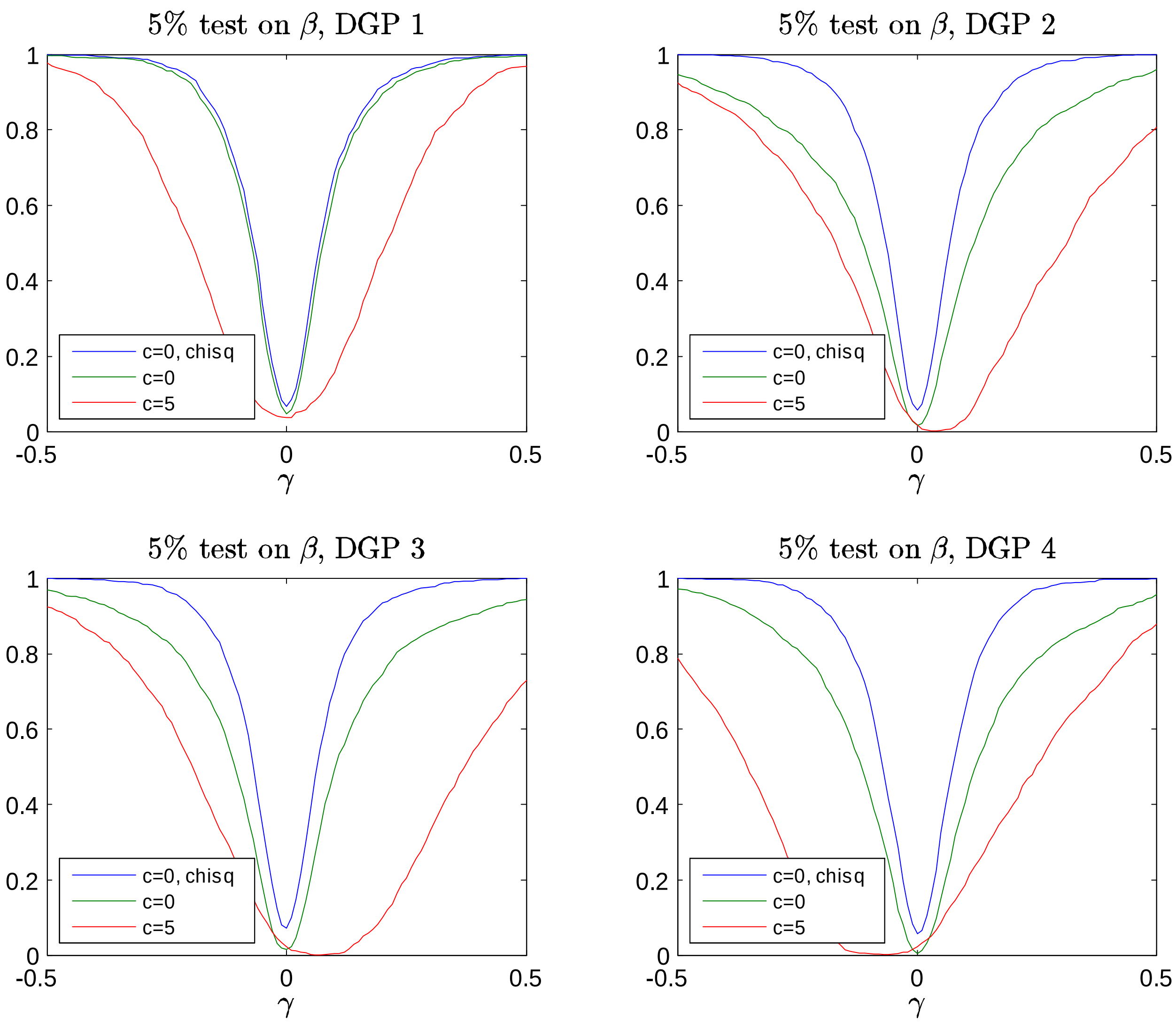

Figure 7: Rejection frequency of the test Q_β for $\gamma = 0$ using the $\chi^2(1)_{0.95}$ quantile (unadjusted) and the adjusted quantile in (19) for $\xi = 95\%$ and $\eta = 5\%$, as a function of γ for different values of c; $N = 1000$ simulations of $T = 100$ observations from the DGPs in Table 1.