European University Institute

DEPARTMENT OF ECONOMICS

EUI Working Papers

ECO 2009/42
DEPARTMENT OF ECONOMICS

GENERALIZED LEAST SQUARES ESTIMATION FOR COINTEGRATION PARAMETERS UNDER CONDITIONAL HETEROSKEDASTICITY

Helmut Herwartz and Helmut Luetkepohl

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE
DEPARTMENT OF ECONOMICS

Generalized Least Squares Estimation for Cointegration Parameters Under Conditional Heteroskedasticity

HELMUT HERWARTZ and HELMUT LUETKEPOHL

EUI Working Paper **ECO** 2009/42

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

December 8, 2009

Generalized Least Squares Estimation for Cointegration Parameters Under Conditional Heteroskedasticity

by

Helmut Herwartz[1]
Institut für Statistik und Ökonometrie, Christian–Albrechts–Universität zu Kiel,
Ohlshausenstr. 40, D–24098 Kiel, GERMANY,
email: h.herwartz@stat-econ.uni-kiel.de

and

Helmut Lütkepohl
Department of Economics, European University Institute,
Via della Piazzuola 43, I–50133 Firenze, ITALY,
email: helmut.luetkepohl@eui.eu

Abstract

In the presence of generalized conditional heteroscedasticity (GARCH) in the residuals of a vector error correction model (VECM), maximum likelihood (ML) estimation of the cointegration parameters has been shown to be efficient. On the other hand, full ML estimation of VECMs with GARCH residuals is computationally difficult and may not be feasible for larger models. Moreover, ML estimation of VECMs with independently identically distributed residuals is known to have potentially poor small sample properties and this problem also persists when there are GARCH residuals. A further disadvantage of the ML estimator is its sensitivity to misspecification of the GARCH process. We propose a feasible generalized least squares estimator which addresses all these problems. It is easy to compute and has superior small sample properties in the presence of GARCH residuals.

Keywords: Vector autoregressive process, vector error correction model, cointegration, reduced rank estimation, maximum likelihood estimation, multivariate GARCH

JEL classification: C32

[1]Much of the research for this study was conducted while the first author was a Fernand Braudel Fellow in the Department of Economics at the EUI.

1 Introduction

In vector autoregressive (VAR) models with cointegrated variables the parameters of the cointegration relations are often of central importance for interpreting the empirical results. Despite the superconsistency of standard estimators for these parameters, their small sample properties are often poor. Therefore asymptotic results due to Seo (2007) are of great importance. He shows that estimation efficiency can be improved by taking into account generalized autoregressive conditionally heteroskedastic (GARCH) residuals.

In practice, the Johansen (1995) reduced rank (RR) maximum likelihood (ML) approach is the most popular method for estimating the cointegration parameters in a vector error correction model (VECM) setup of a VAR. It was derived under the assumption of independently identically distributed (i.i.d.) Gaussian residuals. Even if the residual distribution is non-Gaussian, the estimator has good asymptotic properties under general conditions. Seo (2007) has shown, however, that the estimator is generally inefficient for processes with GARCH residuals. In that case, using a full ML procedure which takes into account the GARCH residuals is efficient. He also presents simulation evidence of efficiency gains in finite samples of size $T = 250$.

Full ML estimation of VECMs with GARCH residuals has at least three major drawbacks, however. First, computation of the estimates is quite demanding and may not even be feasible for larger models with a moderate number of variables and a realistic number of lags. Second, in small samples the ML estimator for VECMs with i.i.d. residuals is known to generate occasional outliers and therefore may yield quite distorted estimates of the cointegration parameters (e.g., Brüggemann & Lütkepohl (2005)). We show that a similar problem can also arise for the full ML estimator when GARCH residuals are accounted for. Third, in practice the precise GARCH structure of the residuals is unknown and ML may not be robust to misspecification of the GARCH process. We propose a feasible generalized least squares (GLS) estimator which takes care of all three problems. It is easy to compute even for large models and has small sample properties superior to full ML for a range of models with features that are typical in empirical studies. It is also more robust to misspecification of the GARCH structure than ML.

In the next section the model setup is presented and the GLS estimator which allows for GARCH residuals is derived in Section 3. A small sample Monte Carlo comparison of the GLS and ML estimators is provided in Section 4. Conclusions are drawn in Section 5.

The following notation is used throughout. The difference operator is denoted by Δ, that

is, for a stochastic process or time series y_t, $\Delta y_t = y_t - y_{t-1}$. The normal distribution with mean (vector) μ and variance (covariance matrix) Σ is signified as $\mathcal{N}(\mu, \Sigma)$. For a symmetric, positive definite matrix Σ, $\Sigma^{-1/2}$ denotes the inverse of the symmetric square root matrix, $|A|$ is the determinant of a square matrix A, I_K is the $(K \times K)$ identity matrix, $\otimes$ signifies the Kronecker product and vec is the vectorization operator which stacks the columns of a matrix in a column vector. An indicator function, denoted $\mathbf{I}(\cdot)$, is one when the conditions in parentheses are satisfied and zero otherwise.

The following abbreviations are used: VAR for vector autoregressive, VECM for vector error correction model, GARCH for generalized autoregressive conditional heteroskedasticity, i.i.d. for independently identically distributed, ML for maximum likelihood, GLS for generalized least squares, OLS for ordinary least squares, RR for reduced rank, CCC for constant conditional correlation, DGP for data generating process, RMSE for root mean squared error, RRMSE for relative root mean squared error, MAE for mean absolute error and RMAE for relative mean absolute error.

2 The Model

Suppose the K-dimensional VAR process y_t has the VECM representation

$$\begin{aligned} \Delta y_t &= \alpha(\beta' y_{t-1} + \delta^{co\prime} d_{t-1}^{co}) + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + C d_t^s + u_t \\ &= \alpha \beta^{*\prime} y_{t-1}^* + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + C d_t^s + u_t. \end{aligned} \tag{2.1}$$

Here α and β are $(K \times r)$ matrices of rank r associated with the long-run part of the model, Γ_i $(i = 1, \ldots, p-1)$ are $(K \times K)$ coefficient matrices associated with the short-run dynamics, d_t^{co} is a vector of deterministic variables which are included in the cointegration relations with corresponding coefficient matrix $\delta^{co\prime}$. The vector d_t^s includes the remaining deterministic variables with coefficient matrix C. The matrix $\beta^{*\prime} = [\beta' : \delta^{co\prime}]$ is $(r \times K^*)$ and $y_{t-1}^* = [y_{t-1}', d_{t-1}^{co\prime}]'$ is $(K^* \times 1)$ with $K^* = K +$ dimension(d_t^{co}). As in Seo (2007), the error term u_t is assumed to be a vector martingale difference sequence with $E(u_t|\mathcal{F}_{t-1}) = 0$ and $E(u_t u_t'|\mathcal{F}_{t-1}) = \Sigma_t$, where $\mathcal{F}_t$ is the σ-field generated by $y_t, y_{t-1}, \ldots$.

In the following the primary objective is to estimate the cointegration parameters β^*. Because the matrix is not unique we impose just-identifying restrictions such that $\beta^{*\prime} = [I_r : \beta_{(K^*-r)}^{*\prime}]$, that is, the first r rows of β^* constitute an $(r \times r)$ identity matrix. As explained in Lütkepohl (2005, Section 6.3), this normalization is not restrictive from a practical point of view.

3 Estimation

Before the GLS estimator for VECMs with GARCH residuals is presented, full Gaussian ML and ML ignoring GARCH are considered.

3.1 ML Estimation with GARCH Residuals

For a given cointegrating rank r and lag order p, a sample with T observations and p presample values, ML estimation of the VECM (2.1) is theoretically straightforward if $u_t|\mathcal{F}_{t-1} \sim \mathcal{N}(0, \Sigma_t)$. Collecting all parameters in a vector θ, the log-likelihood function for the model from Section 2 is $l(\theta) = T^{-1}\sum_{t=1}^{T} l_t(\theta)$, where $l_t(\theta) = -\frac{1}{2}\log|\Sigma_t(\theta)| - \frac{1}{2}u_t(\theta)'\Sigma_t(\theta)^{-1}u_t(\theta)$ and a constant term has been dropped (see Seo (2007, Eq. (6))). Seo (2007) derives the asymptotic properties of the ML estimators of the cointegration parameters under the assumption that either the model has no deterministic terms ($d_t^{co} = 0, d_t^s = 0$) or it has an intercept term ($d_t^{co} = 0, d_t^s = 1$). Moreover, he assumes the GARCH process to be of the constant conditional correlation (CCC) form (Bollerslev (1990)).

3.2 Reduced Rank Estimator

An alternative estimator of the cointegration parameters can be obtained by a pseudo-ML method which ignores the GARCH structure of u_t. As shown in Johansen (1995), it can be obtained by RR regression. In other words, the estimator may be determined by denoting the residuals from regressing Δy_t and y_{t-1}^* on $\Delta Y_{t-1}' = [\Delta y_{t-1}', \ldots, \Delta y_{t-p+1}', d_t^{s\prime}]$ by R_{0t} and R_{1t}, respectively, defining $S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R_{jt}'$ $(i, j = 0, 1)$, and solving the generalized eigenvalue problem $|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0$. Let $\lambda_1 \geq \cdots \geq \lambda_{K^*}$ be the ordered eigenvalues with corresponding matrix of eigenvectors $B = [b_1, \ldots, b_{K^*}]$ satisfying $\lambda_i S_{11}b_i = S_{01}'S_{00}^{-1}S_{01}b_i$ and normalized such that $B'S_{11}B = I_{K^*}$. An estimator of β^* is given by $\tilde{\beta}^* = [b_1, \ldots, b_r]$. Post-multiplying by the inverse of the first r rows of $\tilde{\beta}^*$ gives an RR estimator $\tilde{\beta}_{RR}^{*\prime} = [I_r : \tilde{\beta}_{(K^*-r)}^{*\prime}]$. This estimator is the Gaussian ML estimator if the residuals are i.i.d..

It turns out that the asymptotic distribution of the RR estimator is invariant to conditional heteroskedasticity of the form considered here (see Seo (2007, p. 78)). However, using full ML with allowance for GARCH residuals will result in more efficient estimators if the true GARCH structure is considered. Obviously, assuming a known GARCH structure is a rather strong assumption and it is worth exploring the robustness of the full ML estimator with respect to a violation of this assumption. We will do so in Section 4.

3.3 A Feasible GLS Estimator

A GLS estimator for the cointegration parameters can be derived by premultiplying the VECM (2.1) by $\Sigma_t^{-1/2}$ and using standard rules for the vec operator and the Kronecker product,

$$\Sigma_t^{-1/2}\Delta y_t = \Sigma_t^{-1/2}\alpha\beta^{*\prime}y_{t-1}^* + (\Delta Y_{t-1}'\otimes\Sigma_t^{-1/2})\gamma + \Sigma_t^{-1/2}u_t, \tag{3.1}$$

where $\gamma = \text{vec}[\Gamma_1,\ldots,\Gamma_{p-1},C]$. Replacing γ by its GLS estimator and rearranging terms gives

$$G_{0t} = G_{1t}\text{vec}(\beta^{*\prime}) + \tilde{u}_t, \tag{3.2}$$

where

$$G_{0t} = \Sigma_t^{-1/2}\Delta y_t - (\Delta Y_{t-1}'\otimes\Sigma_t^{-1/2})\left[\sum_{t=1}^{T}\Delta Y_{t-1}\Delta Y_{t-1}'\otimes\Sigma_t^{-1}\right]^{-1}(\Delta Y_{t-1}\otimes\Sigma_t^{-1})\Delta y_t,$$

is a $(K\times 1)$ vector,

$$G_{1t} = (y_{t-1}^{*\prime}\otimes\Sigma_t^{-1/2}\alpha) - (\Delta Y_{t-1}'\otimes\Sigma_t^{-1/2})\left[\sum_{t=1}^{T}\Delta Y_{t-1}\Delta Y_{t-1}'\otimes\Sigma_t^{-1}\right]^{-1}(\Delta Y_{t-1}y_{t-1}^{*\prime}\otimes\Sigma_t^{-1/2}\alpha)$$

is a $(K\times rK^*)$ matrix and $\tilde{u}_t$ is the corresponding error term. Hence, denoting by $G_{1t}^{(1)}$ and $G_{1t}^{(2)}$ the first r^2 and last $r(K^*-r)$ columns of G_{1t}, respectively, and using that $\beta^{*\prime} = [I_r : \beta_{(K^*-r)}^{*\prime}]$, the GLS estimator for $\text{vec}(\beta_{(K^*-r)}^{*\prime})$ can be expressed as

$$\text{vec}(\breve{\beta}_{(K^*-r)}^{*\prime}) = \left(\sum_{t=1}^{T}G_{1t}^{(2)\prime}G_{1t}^{(2)}\right)^{-1}\sum_{t=1}^{T}G_{1t}^{(2)\prime}\left(G_{0t} - G_{1t}^{(1)}\text{vec}(I_r)\right). \tag{3.3}$$

Notice that this estimator is the Gaussian ML estimator of $\beta_{(K^*-r)}^*$ for given α and Σ_t. In other words, if the latter two quantities are replaced by their ML estimators, the GLS estimator is identical to the ML estimator. Given the practical difficulties in computing the ML estimator in the presence of GARCH residuals, we propose the following two-step procedure for a feasible GLS estimator.

Step 1. Estimate the parameters in the model

$$\Delta y_t = \Pi y_{t-1}^* + \Gamma_1\Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + Cd_t^s + u_t$$

by OLS and denote the residuals by $\hat{u}_t$. Then estimate the GARCH parameters, ψ say, from a pseudo ML estimation based on maximizing $\hat{l}(\psi) = T^{-1}\sum_{t=1}^{T}\hat{l}_t(\psi)$, where

$$\hat{l}_t(\psi) = -\frac{1}{2}\log|\Sigma_t(\psi)| - \frac{1}{2}\hat{u}_t'\Sigma_t(\psi)^{-1}\hat{u}_t.$$

Let the estimate be $\hat{\psi}$ and define $\widehat{\Sigma}_t = \Sigma_t(\hat{\psi})$. □

Step 2. Use the $\widehat{\Sigma}_t$ from Step 1 to estimate $\pi = \text{vec}(\Pi) = \text{vec}(\alpha\beta^{*\prime})$ and γ by OLS from the transformed model

$$\widehat{\Sigma}_t^{-1/2}\Delta y_t = (y_{t-1}^{*\prime} \otimes \widehat{\Sigma}_t^{-1/2})\pi + (\Delta Y_{t-1}^{\prime} \otimes \widehat{\Sigma}_t^{-1/2})\gamma + \tilde{v}_t,$$

rearrange the elements of the estimator $\hat{\pi}$ in the matrix $\widehat{\Pi}$ and use the first r columns of $\widehat{\Pi}$ as an estimator $\hat{\alpha}$ of α because the first r columns of Π are equal to α. Finally, use $\hat{\alpha}$ and $\widehat{\Sigma}_t$ in G_{0t} and G_{1t} in place of α and Σ_t, respectively, to compute a feasible GLS estimator of $\text{vec}(\beta_{(K^*-r)}^{*\prime})$ as in (3.3). The feasible GLS estimator is denoted by $\text{vec}(\breve{\breve{\beta}}_{(K^*-r)}^{*\prime})$. □

Although estimation of the GARCH parameters in Step 1 requires numerical optimization of $\hat{l}(\psi)$, this optimization problem involves fewer parameters than optimization of the full likelihood $l(\theta)$. Hence, the numerical evaluation of the GLS estimator is easy compared with full ML because all other computations are based on closed form expressions.

If there are no short-term dynamics, that is,

$$\Delta y_t = \alpha\beta^{*\prime} y_{t-1}^* + u_t, \tag{3.4}$$

the model can be written in the form

$$\Sigma_t^{-1/2}(\Delta y_t - \alpha y_{t-1}^{*(1)}) = (y_{t-1}^{*(2)\prime} \otimes \Sigma_t^{-1/2}\alpha)\text{vec}(\beta_{(K^*-r)}^{*\prime}) + v_t, \tag{3.5}$$

where the normalization $\beta^{*\prime} = [I_r : \beta_{(K^*-r)}^{*\prime}]$ has been used and $v_t = \Sigma_t^{-1/2}u_t$. Furthermore, $y_{t-1}^{*(1)}$ and $y_{t-1}^{*(2)}$ consist of the first r and last $K^* - r$ components of y_{t-1}^*, respectively. For given α and Σ_t, the GLS estimator of $\text{vec}(\beta_{(K^*-r)}^{*\prime})$ in this model simplifies to

$$\text{vec}(\breve{\beta}_{(K^*-r)}^{*\prime}) = \left[\sum_{t=1}^{T} y_{t-1}^{*(2)} y_{t-1}^{*(2)\prime} \otimes \alpha'\Sigma_t^{-1}\alpha\right]^{-1} \sum_{t=1}^{T}(y_{t-1}^{*(2)} \otimes \alpha'\Sigma_t^{-1})(\Delta y_t - \alpha y_{t-1}^{*(1)}). \tag{3.6}$$

These considerations suggest another feasible GLS estimator which does not reestimate π and γ in Step 2 but uses the OLS estimates from Step 1 instead. It may be based on the concentrated model

$$R_{0t} = \alpha\beta^{*\prime} R_{1t} + \tilde{u}_t, \tag{3.7}$$

where R_{0t} and R_{1t} are the residual series from the RR regression discussed in Section 3.2 and $\tilde{u}_t$ is the corresponding error term. Thus, the GLS estimator is obtained by replacing Δy_t

and y^*_{t-1} in (3.6) by R_{0t} and R_{1t}, respectively. Using this setup, it is straightforward to show that, if there is no conditional heteroskedasticity so that $\Sigma_t = \Sigma_u$, the GLS estimator reduces precisely to the estimator considered by Ahn & Reinsel (1990), Reinsel (1993, Chapter 6), Saikkonen (1992) and Brüggemann & Lütkepohl (2005). In the following the estimator which assumes no GARCH, $\Sigma_t = \Sigma_u$, will be denoted by GLS1, the estimator based on (3.6) and (3.7), which does not reestimate α and γ in Step 2, is abbreviated as GLS2 and the full two-step GLS estimator is signified as GLS3.

As usual, an estimator of the covariance matrix of our GLS estimator may be obtained as

$$\left(\sum_{t=1}^{T} G_{1t}^{(2)\prime} G_{1t}^{(2)} \right)^{-1}, \tag{3.8}$$

where again α and Σ_t are replaced by the estimators $\hat{\alpha}$ and $\widehat{\Sigma}_t$, respectively, from Steps 1 and 2. For the special case model (3.4) the estimator simplifies to

$$\left[\sum_{t=1}^{T} y_{t-1}^{*(2)} y_{t-1}^{*(2)\prime} \otimes \hat{\alpha}' \widehat{\Sigma}_t^{-1} \hat{\alpha} \right]^{-1}. \tag{3.9}$$

This estimator is used in computing t-ratios for GLS2 and GLS3 in the simulation experiment reported in the next section. In (3.9), $\hat{\alpha}$ is the OLS estimator for GLS2 and the GLS estimator from Step 2 for GLS3.

4 Monte Carlo Comparison

In the following Monte Carlo experiments we compare the five estimators based on ML without allowance for GARCH (RR), ML with GARCH (ML), GLS1, GLS2 and GLS3. The Monte Carlo design is described in Section 4.1 and the main simulation results are summarized in Section 4.2.

4.1 Monte Carlo Design

Our Monte Carlo experiments use the same bivariate model setup as Seo (2007) but replace some of the parameters by values more common in applied work. Thus, the model is

$$\Delta y_t = \alpha \beta' y_{t-1} + u_t,$$

where $\alpha = (\alpha_1, 0)'$, $\beta = (1, \beta_2)'$ with $\beta_2 = -1$, and u_t is a CCC-GARCH process. More precisely, the conditional covariance, $E(u_t u_t' | \mathcal{F}_{t-1}) = \Sigma_t$, is based on the orthogonalized

Table 1: Parameter Values Used in Monte Carlo Simulations

DGP	α_1	β_2	λ	ω_j	ψ_j	ϕ_j
1	-1.0	-1.0	0	1	0	0
2	-1.0	-1.0	-0.5	1	0.25	0.70
3	-1.0	-1.0	0.5	0.05	0.05	0.90
4	-1.0	-1.0	0.5	0.05	0.10	0.85
5	-1.0	-1.0	0.5	$\underline{\omega}_j$, $\overline{\omega}_j$	0.05	0.90
6	-1.0	-1.0		BEKK		
7	-0.1	-1.0	0	1	0	0
8	-0.1	-1.0	-0.5	1	0.25	0.70
9	-0.1	-1.0	0.5	0.05	0.05	0.90
10	-0.1	-1.0	0.5	0.05	0.10	0.85
11	-0.1	-1.0	0.5	$\underline{\omega}_j$, $\overline{\omega}_j$	0.05	0.90
12	-0.1	-1.0		BEKK		

process e_t defined such that $E[e_t e_t'|\mathcal{F}_{t-1}] = \Omega_t$ with $\Omega_t = \text{diag}(\sigma_{1t}^2, \sigma_{2t}^2)$ and $\sigma_{j,t}^2 = \omega_j + \psi_j e_{j,t-1}^2 + \phi_j \sigma_{j,t-1}^2$, where $e_{jt} = \sigma_{jt}\xi_{jt}$, $\xi_{jt} \sim \mathcal{N}(0,1)$. The relation between u_t and e_t is given by $u_t = Le_t$, where

$$L = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}.$$

Thus, λ determines the correlation between the two components of u_t.

In all the simulations the cointegration parameter $\beta_2 = -1$. This value is in line with typical cointegration parameters in economic models and was also used by Seo (2007). The values of the parameters α, λ, ω_j, ψ_j, ϕ_j, used in the simulations are summarized in Table 1. The first two processes were also considered by Seo (2007) in a simulation study aimed at demonstrating the virtues of ML estimation with GARCH residuals. The first one has no GARCH so that it is suitable for investigating possible estimation efficiency losses due to assuming GARCH when there is none. In all models considered by Seo, $\alpha_1 = -1$. Therefore we also use this parameter value for the first six processes in the simulation comparison. Such strong error correction dynamics make estimation of the cointegration parameter potentially easier. It is not very realistic from a practical point of view because in empirical studies loading coefficients much closer to zero are common. Therefore we use a more realistic value of α_1 in DGPs 7 - 12. The choice of $\alpha_2 = 0$ in all DGPs implies that the second variable is weakly exogenous. This situation is not uncommon in practice and was also used in most of Seo's DGPs.

We also use a range of GARCH specifications different from those employed by Seo. In most of his DGPs the correlation parameter λ is zero implying that the components of u_t follow independent univariate processes which is not of particular interest in a vector GARCH context. The only exception is a process with $\lambda = -0.5$ in one of Seo's DGPs. Again this is a somewhat unrealistic case because it implies that the correlation among the two residual components is negative. Hence, we choose some DGPs with positive values of λ.

As in most of Seo's DGPs, we also choose identical values for the GARCH parameters ω_j, ψ_j, ϕ_j of the two residual components. Again we select complementarily parameter settings that are more in line with empirical analyses of conditionally heteroskedastic processes. Considering higher frequency data, daily or weekly say, GARCH parameters ψ_j tend to be closer to zero while ϕ_j is often close to unity. To mention a particular framework where cointegration features and second order moment characteristics at higher frequency are core model ingredients one could think of term structure modelling/monitoring or price discovery in financial markets (see Diebold & Li (2006) and Hasbrouck (1995)).

To ensure that the simulation results for DGPs 3-6 and 9-12 are unaffected by the unconditional variances of the variables, the intercept term of the GARCH equation is selected such that the unconditional variance is unity. DGPs 5 and 11 are characterized by a shift of the unconditional variance such that after one fourth of the generated time span the unconditional variance shifts by a factor of 4. Accordingly, GARCH intercept parameters are chosen as

$$\omega_{jt} = \underline{\omega}_j \mathbf{I}(t \leq T/4) + \overline{\omega}_j \mathbf{I}(t > T/4),$$

with $\underline{\omega}_j = 0.05/3.25 \approx 0.0154$, $\overline{\omega}_j = 0.05 \times 4/3.25 \approx 0.0615$.

Finally, we consider GARCH processes which do not have a CCC structure. They allow us to determine the effects of other types of misspecifying the residual process. DGPs 6 and 12 are so-called BEKK models (Engle & Kroner (1995)), i.e.,

$$\Sigma_t = \Gamma_0 \Gamma_0' + D u_{t-1} u_{t-1}' D' + F \Sigma_{t-1} F' \tag{4.1}$$

with

$$D = \begin{pmatrix} 0.229 & -0.173 \\ 0.005 & 0.174 \end{pmatrix} \quad \text{and} \quad F = \begin{pmatrix} 0.954 & 0.033 \\ 0.008 & 0.981 \end{pmatrix}.$$

The parameter values are estimates for a bivariate residual series obtained from a VAR that is composed of so-called breakeven inflation rates (generated from prices of nominal and

inflation protected long-term government bonds) prevailing in France and the UK, covering the time period 8/3/2000 - 9/30/2005 (see Hafner & Herwartz (2009) for details). The moduli of the eigenvalues of the matrix $(D \otimes D) + (F \otimes F)$ are close to but smaller than one. Hence, the BEKK process has high persistence in second moments but remains covariance stationary. The variables are again normalized to have unit unconditional variances and Γ_0 is chosen accordingly.

For each DGP, $M = 5000$ sets of time series were generated and the cointegration parameter β_2 was estimated by the five methods RR, ML, GLS1, GLS2 and GLS3. ML estimators were determined using first order numerical derivatives (GAUSS 6.0, *gradp*) to implement BHHH optimization and covariance estimation which is justified given the conditional normality of all DGPs considered (Berndt, Hall, Hall & Hausman (1974)). The estimators will be compared on the basis of their root mean squared error (RMSE), mean absolute error (MAE) and rejection frequencies of t-tests of a null hypothesis $H_0 : \beta_2 = -1$ with significance level 5%. Since the finite sample moments of the ML estimator do not exist, it seems sensible to consider the MAE in addition to the RMSE. The rejection frequencies of the t-tests give an impression of the usefulness of the estimators for doing inference regarding the actual parameter value.

4.2 Monte Carlo Results

Simulation results for sample sizes $T = 100$, 250 and 500 are presented in Tables 2 and 3 as well as Figures 1 - 4. In the tables RMSEs and MAEs relative to RR are given for all estimators but RR and denoted by RRMSE and RMAE, respectively. The actual RMSEs and MAEs of RR, which are given in boldface in the tables, show that estimation precision increases with the sample size, reflecting consistency of the estimator. In Figures 1, 2 and 3 only relative MAEs are depicted. Figure 4 shows relative rejection frequencies for t-tests based on a nominal significance level of 5%. The following observations emerge from these results.

Overall the GLS estimators have smaller RMSEs and MAEs than the RR and ML estimators. There are, in fact, only two processes where ML is clearly preferable, namely when DGPs 2 or 8 and relatively large samples of sizes $T = 250$ or 500 are used. Recall that both DGPs have somewhat uncommon GARCH parameters. DGP 2 was also considered by Seo who does not report results for sample sizes $T = 100$ or 500. Thus, only for two DGPs with unusual GARCH parameters ML has a sizable lead over GLS when at least a moderately

large sample is considered. In all other cases GLS with GARCH correction (GLS2/GLS3) is superior or roughly at the same level in terms of RMSE and MAE. The GLS estimators are substantially superior to ML if there is no GARCH in the residuals (DGPs 1 and 7). This outcome is obtained especially for larger samples ($T = 250$ and 500). It may be an implication of the lack of identification of the GARCH parameters for this case.

All three GLS estimators are typically rather close together in terms of RMSE and MAE. GLS2 and GLS3 results are in fact almost indistinguishable in the figures for all DGPs and sample sizes. Thus, reestimating α in Step 2 of the GLS procedure does not make much difference for estimation precision. Only in rare occasions sizable gains are obtained by accounting for GARCH. This behaviour prevails even in larger samples. For our simple models, the superconsistency of the estimators may well become effective for samples of size $T = 250$ or 500 so that for such sample sizes accounting for specific residual properties is of limited value for the precision of the estimators.

ML performs particularly poorly for smaller samples ($T = 100$) when the loading coefficient is large in absolute value ($\alpha = -1$) and there is no GARCH (DGP 1), realistic GARCH parameters are considered (DGPs 3 and 4) or the GARCH process is misspecified (DGPs 5 and 6). In these cases ML is in fact inferior to all other estimators in terms of RMSE and/or MAE. The very poor performance of ML for DGP 1 even for large samples (see Figure 3) is likely to be driven by the fact that the GARCH parameters are not identified in this case and this affects the estimator for the cointegration parameter if a full likelihood optimization is attempted. The results for DGP 1 illustrate particularly forcefully the lack of robustness of the ML estimator. The relatively poor performance of RR in small samples ($T = 100$) for DGPs with more realistic loading coefficients (see, e.g., DGPs 9 - 12 in Table 3) reflects its heavy-tailed small sample distribution (see Brüggemann & Lütkepohl (2005)).

The rejection frequencies for testing $H_0 : \beta_2 = -1$ by means of t-ratios with 5% nominal significance level vary substantially across DGPs for all the estimators (see Tables 2 and 3 and Figure 4). In some cases the relative rejection frequencies are considerably larger than 0.05, especially when the loading coefficient α has a more realistic value (see, e.g., DGPs 7 - 12). As can be seen in Figure 4, all tests overreject for these processes even for samples as large as $T = 250$. In fact, the rejection frequencies based on the ML estimator for DGP 1 which does not have GARCH residuals, do not even improve with increasing sample size due to the lack of identification of the GARCH parameters (see the results for $T = 500$ in Table 2). Thus, ML is particularly unreliable for inference if the true DGP is unknown. In contrast, the GLS estimators are more robust and the rejection frequencies improve with

increasing sample size for all DGPs.

In summary, these observations imply that the GLS estimators are overall superior to ML in terms of RMSE and MAE if DGPs with realistic parameter values are used. Comparing GLS with and without allowance for GARCH shows that gains from taking GARCH into account are limited, in particular, if the true model structure is unknown. Inference based on any of the estimators is likely to be distorted in samples of realistic size. ML with allowance for GARCH may produce very unreliable estimates especially if the assumed GARCH process is not the true one. Obviously, the latter situation is common in practice.

4.3 Other Simulation Experiments

We have also considered other estimators and simulation setups for which we do not report detailed results because they do not add much to the conclusions we can draw from the results in Tables 2 and 3 and Figures 1 - 4. They may still be worth summarizing here.

First of all we have considered even smaller samples of size $T = 50$. For such small samples RR produces extremely poor RMSEs and MAEs especially when $\alpha_1 = -0.1$. This result is in line with the simulations reported by Brüggemann & Lütkepohl (2005). It is due to the heavy-tailed small sample distribution of the RR estimator (for a theoretical analysis see Phillips (1994)). Clearly, one would not expect the problem to disappear when there is GARCH in the residuals. It is not surprising that the problem also arises to some extent for the ML estimator which accounts for GARCH, although to a lesser extent. We do not report detailed results because for samples as small as $T = 50$, GARCH residuals are typically not viewed as a major problem in applied work.

Instead of using the RR estimator we have computed the ML estimator without allowance for GARCH residuals by a numerical algorithm similar to the one used in the GARCH ML procedure. This was done primarily to check our algorithms. Although there were slight deviations in the estimates, the results were very similar, in particular for larger samples. In small samples ($T = 50$) numerical differences were found between RR and the ML algorithm without GARCH.

To further improve the robustness of the t-tests based on the GLS estimator we have also considered a "heteroskedasticity consistent" covariance estimator of the asymptotic distribution of GLS2 and GLS3 of the form

$$\left[\sum_{t=1}^{T} y_{t-1}^{*(2)} y_{t-1}^{*(2)\prime} \otimes \hat{\alpha}'\widehat{\Sigma}_t^{-1}\hat{\alpha}\right]^{-1} \left\{\sum_{t=1}^{T} y_{t-1}^{*(2)} y_{t-1}^{*(2)\prime} \otimes \hat{\alpha}'\widehat{\Sigma}_t^{-1}\hat{\tilde{u}}_t\hat{\tilde{u}}_t'\widehat{\Sigma}_t^{-1}\hat{\alpha}\right\} \left[\sum_{t=1}^{T} y_{t-1}^{*(2)} y_{t-1}^{*(2)\prime} \otimes \hat{\alpha}'\widehat{\Sigma}_t^{-1}\hat{\alpha}\right]^{-1}.$$

Here $\hat{\tilde{u}}_t$ denotes a residual based on the corresponding GLS estimator. Unfortunately, precision gains for the t-tests were rather limited, especially for the misspecified GARCH models where one might have expected more substantial improvements.

5 Conclusions

Despite the superconsistency of standard estimators for the cointegration parameters in a VAR model, the small sample properties are often poor. This state of affairs is particularly problematic because the cointegration parameters are often of central interest for interpreting empirical findings. Therefore attempts have been made to improve the small sample estimation efficiency. In particular, it was proposed recently to use ML estimation when the residuals have GARCH properties. In this study we use a Monte Carlo experiment to show that ML is problematic in the present context because it may produce very unreliable estimates when realistic parameter values for the conditional first and second moments of the model are considered. We propose feasible GLS estimators which produce more efficient estimates for the cointegration parameters in terms of RMSE and MAE than ML in a simple model setup. They are also more robust than ML to misspecification of the GARCH process. Moreover, the GLS estimators have the advantage of being much easier to compute than the full ML estimator accounting for GARCH residuals. Overall, the efficiency gains from accounting for GARCH are rather limited, however.

In the simulation study we have assumed that the cointegrating rank and the lag order of the VAR process are known. Thereby we have been able to focus attention on the differences between the estimators under ideal conditions. Clearly, this is a simplification which is hardly realistic. Even though unit root and cointegration tests which account for heteroskedasticity have been proposed, it is by no means clear that they will necessarily lead to a correctly specified cointegrating rank in practice. Moreover, the models used in the Monte Carlo study are particularly simple in that there is only one cointegration parameter to estimate. A simple model setup is a concession to the ML estimator which is computationally problematic for more complicated models. However, even for our simple model setup ML does not work well relative to GLS. We have no reason to believe that the relative performance of ML improves in more complicated setups. Thus, we conclude that if one wants to account for GARCH in the residuals of a VAR, our GLS estimators for the cointegration parameters are preferable to ML. In fact, GLS2, the simpler version based on OLS estimates of the short-run parameters, can be recommended because it is very easy to compute and performs about as well as the

full two-step estimator.

References

Ahn, S. K. & Reinsel, G. C. (1990). Estimation of partially nonstationary multivariate autoregressive models, *Journal of the American Statistical Association* **85**: 813–823.

Berndt, E. K., Hall, B. H., Hall, R. E. & Hausman, J. A. (1974). Estimation and inference in nonlinear structural models, *Annals of Economic and Social Measurement* **3/4**: 653–665.

Bollerslev, T. (1990). Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model, *Review of Economics and Statistics* **72**: 498–505.

Brüggemann, R. & Lütkepohl, H. (2005). Practical problems with reduced-rank ML estimators for cointegration parameters and a simple alternative, *Oxford Bulletin of Economics and Statistics* **67**: 673–690.

Diebold, F. X. & Li, C. (2006). Forecasting the term structure of government bond yields, *Journal of Econometrics* **130**: 337–364.

Engle, R. F. & Kroner, K. F. (1995). Multivariate simultaneous generalized GARCH, *Econometric Theory* **11**: 122–150.

Hafner, C. M. & Herwartz, H. (2009). Testing for linear vector autoregressive dynamics under multivariate generalized autoregressive heteroskedasticity, *Statistica Neerlandica* **63**: 294–323.

Hasbrouck, J. (1995). One security, many markets, determining the contributions to price discovery, *Journal of Finance* **50**: 1175–1199.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Phillips, P. C. B. (1994). Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models, *Econometrica* **62**: 73–93.

Reinsel, G. C. (1993). *Elements of Multivariate Time Series Analysis*, Springer-Verlag, New York.

Saikkonen, P. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation, *Econometric Theory* **8**: 1–27.

Seo, B. (2007). Asymptotic distribution of the cointegrating vector estimator in error correction models with conditional heteroskedasticity, *Journal of Econometrics* **137**: 68–111.

Table 2: Simulation Results for Estimators of β_2 Based on DGPs 1 - 6

DGP	Estim.	RRMSE	RMAE	5%	RRMSE	RMAE	5%	RRMSE	RMAE	5%
		$T = 100$			$T = 250$			$T = 500$		
1	RR	**0.0195**	**0.0131**	0.0666	**0.0085**	**0.0058**	0.0546	**0.0045**	**0.0031**	0.0518
	ML	1.2103	1.2894	0.1332	1.3194	1.4176	0.1766	1.4074	1.4626	0.1944
	GLS1	0.9986	0.9988	0.0670	0.9992	0.9998	0.0550	1.0001	1.0001	0.0514
	GLS2	1.0059	1.0037	0.0696	1.0012	1.0024	0.0558	1.0009	1.0001	0.0552
	GLS3	1.0053	1.0034	0.0688	1.0010	1.0025	0.0556	1.0011	1.0003	0.0552
2	RR	**0.0155**	**0.0102**	0.0634	**0.0067**	**0.0045**	0.0538	**0.0036**	**0.0025**	0.0562
	ML	0.9732	0.9895	0.0804	0.8322	0.8289	0.0514	0.7631	0.7744	0.0538
	GLS1	0.9978	0.9988	0.0634	1.0008	1.0003	0.0538	1.0007	1.0002	0.0560
	GLS2	0.9147	0.9282	0.0710	0.8761	0.8700	0.0606	0.8200	0.8278	0.0640
	GLS3	0.9108	0.9260	0.0712	0.8740	0.8669	0.0610	0.8173	0.8268	0.0622
3	RR	**0.0173**	**0.0118**	0.0662	**0.0075**	**0.0052**	0.0598	**0.0040**	**0.0028**	0.0498
	ML	1.1358	1.1474	0.0888	1.0682	1.0734	0.0758	1.0133	1.0037	0.0580
	GLS1	0.9992	0.9990	0.0660	0.9979	0.9992	0.0598	0.9998	0.9998	0.0496
	GLS2	1.0017	1.0049	0.0734	0.9976	0.9989	0.0604	0.9944	0.9879	0.0568
	GLS3	1.0018	1.0045	0.0722	0.9972	0.9987	0.0606	0.9951	0.9882	0.0570
4	RR	**0.0173**	**0.0118**	0.0684	**0.0074**	**0.0052**	0.0584	**0.0040**	**0.0028**	0.0520
	ML	1.0770	1.0844	0.0820	0.9871	0.9864	0.0600	0.9492	0.9390	0.0552
	GLS1	0.9991	0.9985	0.0684	0.9984	0.9993	0.0586	0.9999	0.9998	0.0516
	GLS2	0.9872	0.9932	0.0752	0.9771	0.9771	0.0596	0.9645	0.9534	0.0584
	GLS3	0.9863	0.9925	0.0742	0.9768	0.9767	0.0582	0.9653	0.9536	0.0592
5	RR	**0.0195**	**0.0138**	0.0678	**0.0086**	**0.0061**	0.0724	**0.0046**	**0.0032**	0.0640
	ML	1.1435	1.1628	0.0934	0.9929	0.9916	0.0756	0.9427	0.9483	0.0716
	GLS1	0.9991	0.9994	0.0676	0.9984	0.9992	0.0726	0.9998	0.9999	0.0638
	GLS2	1.0001	1.0048	0.0716	0.9879	0.9836	0.0706	0.9688	0.9693	0.0630
	GLS3	1.0004	1.0052	0.0708	0.9877	0.9835	0.0710	0.9701	0.9702	0.0624
6	RR	**0.0183**	**0.0125**	0.0646	**0.0082**	**0.0056**	0.0564	**0.0043**	**0.0029**	0.0534
	ML	1.1444	1.1881	0.1060	1.0777	1.0974	0.0856	1.0192	1.0084	0.0616
	GLS1	0.9984	0.9992	0.0652	0.9974	0.9991	0.0564	0.9995	1.0000	0.0536
	GLS2	0.9859	0.9892	0.0670	0.9722	0.9785	0.0552	0.9746	0.9735	0.0514
	GLS3	0.9849	0.9886	0.0660	0.9739	0.9794	0.0556	0.9742	0.9733	0.0510

Note: The numbers listed in boldface for RR are RMSEs and MAEs.

Table 3: Simulation Results for Estimators of β_2 Based on DGPs 7 - 12

DGP	Estim.	RRMSE	RMAE	5%	RRMSE	RMAE	5%	RRMSE	RMAE	5%
		$T = 100$			$T = 250$			$T = 500$		
7	RR	**0.5123**	**0.1786**	0.1918	**0.0986**	**0.0631**	0.0992	**0.0470**	**0.0325**	0.0776
	ML	0.4293	0.8209	0.1546	1.0075	1.0498	0.0978	1.1108	1.0764	0.0960
	GLS1	0.4376	0.8182	0.1774	0.9397	0.9798	0.0976	0.9934	0.9954	0.0792
	GLS2	0.4399	0.8207	0.1816	0.9396	0.9821	0.1002	0.9939	0.9960	0.0804
	GLS3	0.4410	0.8228	0.1832	0.9401	0.9831	0.1000	0.9949	0.9963	0.0802
8	RR	**0.4690**	**0.1437**	0.1598	**0.0783**	**0.0507**	0.0924	**0.0393**	**0.0260**	0.0794
	ML	0.4546	0.8214	0.1016	0.7944	0.8130	0.0646	0.7419	0.7652	0.0608
	GLS1	0.4110	0.8540	0.1560	0.9643	0.9896	0.0942	1.0021	1.0014	0.0792
	GLS2	0.3761	0.7914	0.1890	0.8510	0.8769	0.1132	0.8360	0.8466	0.0966
	GLS3	0.3727	0.7833	0.1666	0.8410	0.8623	0.0982	0.8058	0.8222	0.0814
9	RR	**1.0982**	**0.1855**	0.2148	**0.0993**	**0.0581**	0.1172	**0.0424**	**0.0291**	0.0806
	ML	0.1973	0.6944	0.1594	0.8491	0.9663	0.1054	0.9757	0.9777	0.0760
	GLS1	0.1828	0.7012	0.1890	0.8013	0.9472	0.1122	0.9788	0.9892	0.0804
	GLS2	0.1822	0.7000	0.1938	0.7991	0.9413	0.1148	0.9737	0.9796	0.0874
	GLS3	0.1832	0.7008	0.1938	0.7989	0.9399	0.1104	0.9758	0.9793	0.0866
10	RR	**0.6809**	**0.1779**	0.2106	**0.4941**	**0.0649**	0.1186	**0.0425**	**0.0291**	0.0804
	ML	3.0086	0.9263	0.1584	0.1734	0.8451	0.1046	0.9425	0.9337	0.0786
	GLS1	0.2988	0.7285	0.1864	0.1629	0.8520	0.1144	0.9832	0.9909	0.0826
	GLS2	0.2935	0.7186	0.1956	0.1593	0.8256	0.1218	0.9515	0.9508	0.0926
	GLS3	0.2978	0.7171	0.1910	0.1586	0.8222	0.1146	0.9503	0.9468	0.0864
11	RR	**4.3177**	**0.2970**	0.2240	**0.1054**	**0.0676**	0.1378	**0.0487**	**0.0339**	0.1020
	ML	0.2615	0.5680	0.1592	0.9125	0.9717	0.1162	0.9379	0.9444	0.0930
	GLS1	0.0519	0.5134	0.1974	0.8752	0.9543	0.1348	0.9798	0.9890	0.1008
	GLS2	0.0517	0.5127	0.1948	0.8644	0.9416	0.1254	0.9474	0.9559	0.0946
	GLS3	0.0535	0.5155	0.1958	0.8635	0.9420	0.1230	0.9484	0.9561	0.0914
12	RR	**4.9552**	**0.3134**	0.2010	**0.1037**	**0.0630**	0.1160	**0.0459**	**0.0307**	0.0794
	ML	0.0758	0.4606	0.1520	0.8906	0.9605	0.1006	0.9668	0.9726	0.0708
	GLS1	0.0432	0.4441	0.1822	0.8566	0.9496	0.1142	0.9848	0.9966	0.0778
	GLS2	0.0426	0.4388	0.1866	0.8325	0.9313	0.1138	0.9596	0.9730	0.0792
	GLS3	0.0427	0.4397	0.1850	0.8305	0.9300	0.1114	0.9550	0.9703	0.0754

Note: The numbers listed in boldface for RR are RMSEs and MAEs.

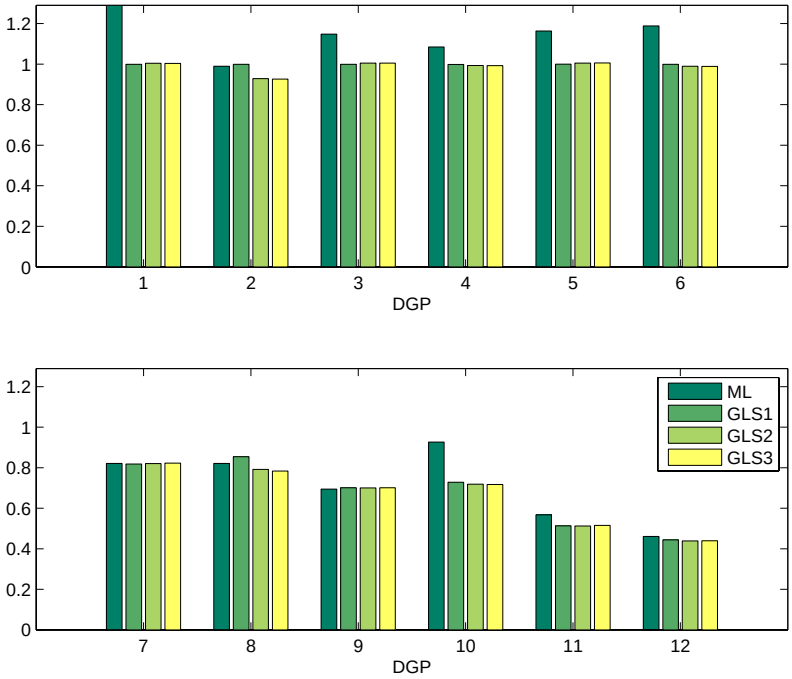

Figure 1: MAEs of estimators relative to RR for sample size $T = 100$.

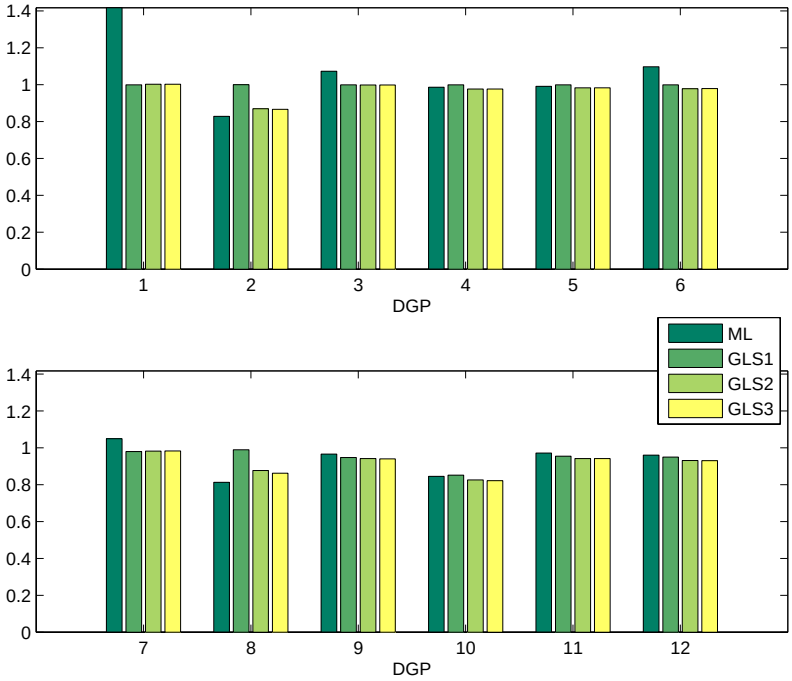

Figure 2: MAEs of estimators relative to RR for sample size $T = 250$.

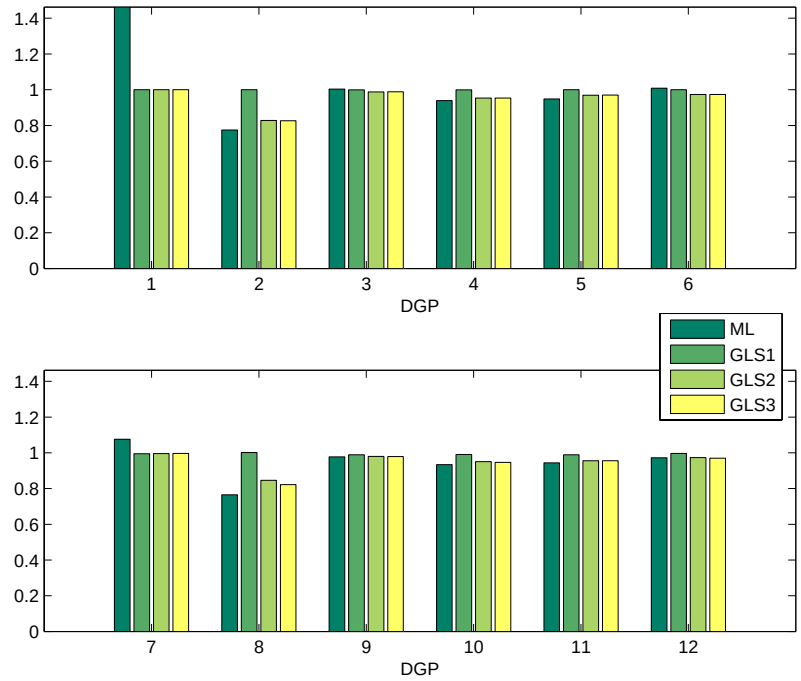

Figure 3: MAEs of estimators relative to RR for sample size $T = 500$.

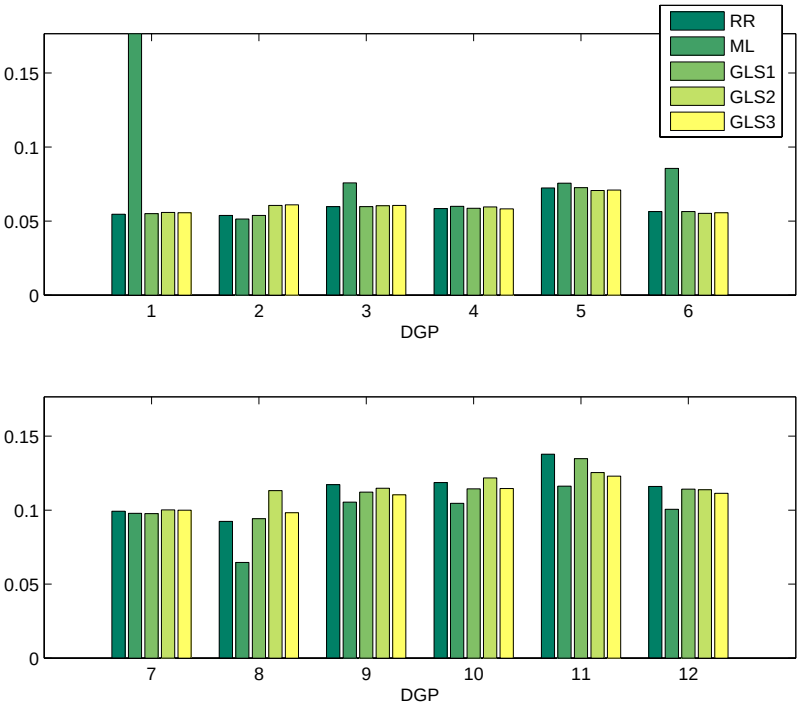

Figure 4: Relative rejection frequencies in 5000 Monte Carlo replications of t-tests with nominal 5% significance level for sample size $T = 250$.