Bootstrap, Jackknife and COLS: bias and mean squared error in estimation of autoregressive models

Gareth Liu-Evans[1] Garry Phillips[2]

[1]University of Liverpool, UK
[2]Cardiff University, UK

Conference in Honour of Andrew Harvey
Oxford, June 29th 2012

Outline

Bias in OLS estimation of autoregressive models

- We compare by Monte Carlo the remaining bias and Mean Squared Error of estimators that aim to correct the OLS bias.
  - Residual bootstrap, Quenoulli half-sample jackknife, and COLS methods
  - AR(1), AR(2), and ARX(1) models with normal and non-normal error terms
  - AR(1) + GARCH(1,1) with Gaussian specification
    - estimators as before but with residual bootstrap replaced by the two-point recursive-design wild bootstrap used in Goncalves and Killian (2004) and Godfrey and Tremayne (2005)

AR(1), AR(2), ARX(1)

$$y_t = \beta + \lambda y_{t-1} + u_t$$
$$y_0 = E[y_t] = \frac{\beta}{1-\lambda}$$
$$u_t \overset{iid}{\sim} N(0,1)\ ,\ \ u_t \overset{iid}{\sim} G^\star(1, 0.5) \text{ or }\ \ u_t \overset{iid}{\sim} P^\star(1, 4.1)$$

$$y_t = \beta + \lambda_1 y_{t-1} + \lambda_2 y_{t-2} + u_t$$
$$y_0 = E[y_t] = \frac{\beta}{1-\lambda_1-\lambda_2},\ \ y_1 = E[y_t] = \frac{\beta}{1-\lambda_1-\lambda_2}.$$

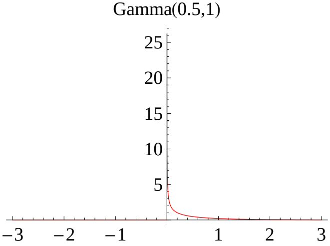
$$y_t = \beta + \lambda y_{t-1} + \delta x_t + u_t, \qquad x_t = \alpha + \rho x_{t-1} + v_t$$
$$y_0 = E[y_t] = \frac{\beta(1-\rho)+\delta\alpha}{(1-\rho)(1-\lambda)}.$$

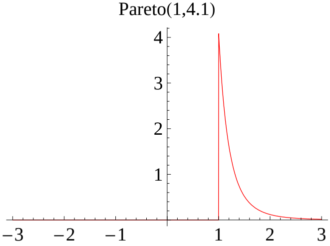
Error moments

Distributions used for the disturbances

	Distribution	μ	σ_u^2	γ_1	γ_2
$N(0,1)$	Standard Normal	0	1	0	0
$G(0.5,1)$	Gamma distribution	0.5	0.707	2.828	12
$P(1,4.1)$	Pareto distribution	1.323	0.451	6.636	786.665

Error moments (continued)





AR(1) + GARCH(1,1)

$$y_t = \beta + \lambda y_{t-1} + u_t,\ \ \sigma_t^2 = \bar{\omega} + \alpha_1 u_{t-1}^2 + \alpha_2 \sigma_{t-1}^2$$
$$y_0 = \beta/(1-\lambda),\ \ \sigma_0^2 = \bar{\omega}/(1-\alpha_1-\alpha_2),\ \ u_0^2 = 0 \qquad (1)$$

The data for this case is generated as follows:

$$\sigma_1^2 = \bar{\omega} + \alpha_2 \sigma_0^2$$
$$\Rightarrow y_1 \ \text{ using a draw from } u_1 \sim N(0, \sigma_1^2)$$
$$\sigma_2^2 = \bar{\omega} + \alpha_1 u_1^2 + \alpha_2 \sigma_1^2$$
$$\Rightarrow y_2 \ \text{ using a draw from } u_2 \sim N(0, \sigma_2^2)$$
$$\sigma_3^2 = \bar{\omega} + \alpha_1 u_2^2 + \alpha_2 \sigma_2^2$$
$$\Rightarrow y_3 \ \text{ using a draw from } u_3 \sim N(0, \sigma_3^2)$$
$$\ldots$$

Residual bootstrap

- For the AR(1) with unknown mean, we generate $y = (y_1, ..., y_T)'$ and $y_{-1} = (y_0, ..., y_{T-1})'$
- OLS is used to obtain $\hat{\beta}$, $\hat{\lambda}$ and $\hat{u} = (\hat{u}_1, ..., \hat{u}_T)'$
- Bootstrap samples $y^\star = (y_1^\star, ..., y_T^\star)'$ and $y_{-1}^\star = (y_0^\star, ..., y_{T-1}^\star)'$ are generated recursively as $y^\star = [i : y_{-1}^\star](\hat{\beta}, \hat{\lambda})' + u^\star$
- The initial value $y_0^\star$ is set as y_0
- The elements of $u^\star = (u_1^\star, ..., u_T^\star)'$ are drawn from $u_t^\star \sim EDF(\hat{u}_t)$
- The EDF assigns probability $\frac{1}{n}$ to each element of $\hat{u}$
- For each bootstrap sample, $y^\star$ is regressed on $[i : y_{-1}^\star]$ to give a bootstrap replicate $\hat{\lambda}_b^\star$

Residual bootstrap

Given $\hat{\lambda}_{\tilde{b}} = \frac{1}{B}\sum_{b=1}^{B} \hat{\lambda}_b^{\star}$ as the mean of the bootstrap replicates and $\hat{\lambda}$ as the OLS estimator of λ, the bootstrap bias-corrected estimator $\hat{\lambda}_b$ is given by

$$\hat{\lambda}_b = 2\hat{\lambda} - \lambda_{\tilde{b}},$$

- The procedure for the AR(2) and ARX(1) is similar.

Quenouille Jackknife

Let $\hat{\lambda}$ be the standard OLS estimator based on all observations, let $\hat{\lambda}_1$ be an OLS estimator that uses only the first half of the sample, and let $\hat{\lambda}_2$ be an OLS estimator that uses only the second half of the sample.
Then the Quenouille jackknife estimator is, following Dhaene et al. (2006),

$$\hat{\lambda}_Q = \frac{T}{T - \lceil T/2 \rceil}\hat{\lambda} - \frac{\lceil T/2 \rceil}{T - \lceil T/2 \rceil}\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{2},$$

which, for even T, is

$$\hat{\lambda}_Q = 2\hat{\lambda} - \left(\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{2}\right).$$

Quenouille Jackknife - intuition/AR(1) example

Kendall (1954) and Marriott and Pope (1954) find the bias in OLS estimation of the Gaussian AR(1) model to be

$$E[\hat{\lambda} - \lambda] = -\frac{1 + 3\lambda}{T} + o(T^{-1}).$$

Therefore, with two even subsamples, we have

$$E[\hat{\lambda}_i - \lambda] = -\frac{1 + 3\lambda}{T/2} + o(T^{-1}), \qquad i = 1 \text{ and } 2,$$

and a substitution of $E[\hat{\lambda}]$, $E[\hat{\lambda}_1]$ and $E[\hat{\lambda}_2]$ into $\hat{\lambda}_Q$ shows that $\hat{\lambda}_Q$ is unbiased to order $O(T^{-1})$.

Higher-order jackknife estimator

- Let λ_Q be the Quenouille jackknife estimator based on the whole sample, and let $\lambda_{1,Q}$ and $\lambda_{2,Q}$ be the Quenouille jackknife estimators based on the individual $T/2$ subsamples.
- Then the following jackknife estimator is unbiased to order $O(T^{-2})$ for the Gaussian autoregressive model with intercept and one lag:

$$\lambda_{QQ} = \frac{4}{3}\lambda_Q - \frac{1}{6}(\lambda_{1,Q} + \lambda_{2,Q}).$$

Higher-order jackknife proof

- In the AR(1) model with intercept and fixed startup, $E[\hat{\lambda}] = \lambda - \frac{1+3\lambda}{T} - \frac{1-3\lambda+9\lambda^2}{(1-\lambda)T^2} + (T^{-2})$.
- From this one can find the expectation of the Quenouille jackknife to order $O(T^{-2})$: $E[\hat{\lambda}_Q] = E[2\hat{\lambda} - \frac{1}{2}(\hat{\lambda}_1 + \hat{\lambda}_2)] = 2(\frac{1-3\lambda+9\lambda^2}{(1-\lambda)T^2}) + o(T^{-2})$.
- Thus, $E[2\hat{\lambda} + \hat{\lambda}_Q] = 3\lambda - 2\frac{1+3\lambda}{T}$, or $E[\frac{2\hat{\lambda}+\hat{\lambda}_Q}{3}] = \lambda - \frac{2}{3}(\frac{1+3\lambda}{T}) + o(T^{-2})$.

Higher-order jackknife proof (continued)

- Let $\lambda^\star = \frac{2\hat{\lambda}+\hat{\lambda}_Q}{3}$, $\lambda_i^\star = \frac{2\hat{\lambda}_i+\hat{\lambda}_{i,Q}}{3}$ for $i = 1, 2$, where $\hat{\lambda}_i$ are the half-sample OLS estimators and $\hat{\lambda}_{i,Q}$ are the Quenouille jackknife estimators based on the *ith* subsample of sample size $T/2$, dividing this into smaller subsamples of size $T/4$.
- Also, let $\lambda_{QQ} = 2\lambda^\star - \frac{1}{2}(\lambda_1^\star + \lambda_2^\star)$.
- Then it follows that $E[\lambda_{QQ}] = \lambda + o(T^{-2})$.
- It can also be shown that the above definition of λ_{QQ} can be rearranged to $\lambda_{QQ} = \frac{4}{3}\lambda_Q - \frac{1}{6}(\lambda_{1,Q} + \lambda_{2,Q})$.

AR(1), bias approximations and COLS

Two examples using the AR(1):

$$E[\hat{\lambda}] = \lambda - \frac{1+3\lambda}{T} + o(T^{-1})$$
$$\Rightarrow \hat{\lambda}_{COLS1} = \hat{\lambda} + \frac{1+3\hat{\lambda}}{T}$$
$$\Rightarrow \tilde{\lambda}_{COLS1} = \frac{T\hat{\lambda}+1}{T-3}.$$

- Both are unbiased to $O(T^{-1})$

AR(1), higher-order moment approximations

$$E[\hat{\lambda}-\lambda] = -\frac{1+3\lambda}{T} - \frac{1-3\lambda+9\lambda^2}{T^2(1-\lambda)} + \frac{1+3\lambda}{T^2}\left[\frac{1}{\sigma^2}\left(\bar{y}_0 - \frac{\beta}{1-\lambda}\right)^2 + \omega^2\right] + o(T^{-2})$$

and

$$E[\hat{\lambda}-\lambda] = -\frac{1+3\lambda}{T} - \frac{1-3\lambda+9\lambda^2}{T^2(1-\lambda)} + \frac{1+3\lambda}{T^2}\left[\frac{1}{\sigma^2}\left(\bar{y}_0 - \frac{\beta}{1-\lambda}\right)^2\right] + \frac{4\gamma_1^2\lambda^2(1-\lambda^2)}{T^2(1-\lambda^3)} + \frac{2\gamma_2\lambda}{T^2} + o(T^{-2}),$$

where $\omega^2\sigma_u^2$ is the variance of y_0.

AR(1), higher-order COLS

These suggest the following COLS2 estimators for the fixed mean-stationary startup:

$$\hat{\lambda}_{COLS2,KP} = \hat{\lambda} + \frac{1 + 3\hat{\lambda}}{T} + \frac{1 - 3\hat{\lambda} + 9\hat{\lambda}^2}{T^2(1 - \hat{\lambda})}$$

and

$$\hat{\lambda}_{COLS2,BAO} = \hat{\lambda} + \frac{1 + 3\hat{\lambda}}{T} + \frac{1 - 3\hat{\lambda} + 9\hat{\lambda}^2}{T^2(1 - \hat{\lambda})} - \frac{4\hat{\gamma}_1^2\hat{\lambda}^2(1 - \hat{\lambda}^2)}{T^2(1 - \hat{\lambda}^3)} - \frac{2\hat{\gamma}_2\hat{\lambda}}{T^2}$$

AR(1), higher-order COLS (continued 1)

- These are not unbiased to order $O(T^{-2})$ though, because the use of $\hat{\lambda}$ in the first correction term $(\frac{1+3\hat{\lambda}}{T})$ creates an additional $O(T^{-2})$ term.
- One method of dealing with this is to add $\frac{3(1+3\hat{\lambda})}{T^2}$.
- A more general way is to use $\hat{\lambda}_{COLS1}$ instead of $\hat{\lambda}$ either everywhere on the right hand sides or just in the $\frac{1+3\hat{\lambda}}{T}$ terms.
- Given the denominators in $\hat{\lambda}_{COLS2,KP}$ and $\hat{\lambda}_{COLS2,BAO}$, it makes sense just to use $\hat{\lambda}_{COLS1}$ in the first term, since $\hat{\lambda}_{COLS1}$ will tend to be closer to unity than $\hat{\lambda}$ when λ is near 1.

AR(1), higher-order COLS (continued 2)

The following are unbiased to order $O(T^{-2})$:

$$\hat{\lambda}_{COLS2,\star,KP} = \hat{\lambda} + \frac{1 + 3\hat{\lambda}_{COLS1}}{T} + \frac{1 - 3\hat{\lambda} + 9\hat{\lambda}^2}{T^2(1 - \hat{\lambda})}$$

and

$$\hat{\lambda}_{COLS2,\star,BAO} = \hat{\lambda} + \frac{1 + 3\hat{\lambda}_{COLS1}}{T} + \frac{1 - 3\hat{\lambda} + 9\hat{\lambda}^2}{T^2(1 - \hat{\lambda})} - \frac{4\hat{\gamma}_1^2\hat{\lambda}^2(1 - \hat{\lambda}^2)}{T^2(1 - \hat{\lambda}^3)} - \frac{2\hat{\gamma}_2\hat{\lambda}}{T^2}.$$

AR(2), bias approximations and COLS

$$E[\hat{\lambda}_1 - \lambda_1] = -\frac{1 + \lambda_1 + \lambda_2}{T} + o\left(T^{-1}\right)$$
$$E[\hat{\lambda}_2 - \lambda_2] = -\frac{2 + 4\lambda_2}{T} + o\left(T^{-1}\right),$$

suggesting

$$\hat{\lambda}_{1,COLS} = \hat{\lambda}_1 + \frac{1 + \hat{\lambda}_1 + \hat{\lambda}_2}{T}$$
$$\hat{\lambda}_{2,COLS} = \hat{\lambda}_2 + \frac{2 + 4\hat{\lambda}_2}{T}.$$

ARX(1), bias approximation

$$E[\hat{\lambda} - \lambda] = \sigma_u^2(-q_{11} tr(Q\bar{Z}'C\bar{Z}) - q_1'\bar{Z}'C\bar{Z}q_1) - 2\sigma_u^4 q_{11}^2 tr(GG'C) + o(T^{-1}),$$

where $\bar{Z} = E[Z] = [\bar{y}_{-1} : X]$ and $\bar{y}_{-1} = E[y_{-1}]$, and where $C = G[0_{T\times 1} : I_T]'$, $D = Z'Z$, $\bar{D} = E[D] = \bar{Z}'\bar{Z} + \sigma_u^2 tr(GG')e_1e_1'$, $Q = \bar{D}^{-1}$, $q_1 = Qe_1$ and $q_{11} = e_1'q_1$. The matrix G, which comes from a decomposition $y_{-1} = \bar{y}_{-1} + Gv$, is $G = \Delta^{-1}[I_T : 0_{T\times 1}]\Omega$, where

$$\Delta = \begin{pmatrix} 1 & 0 & . & . & . & 0 \\ -\lambda & 1 & & & & . \\ 0 & -\lambda & 1 & & & . \\ . & . & . & . & & \\ . & & . & . & . & . \\ 0 & . & . & 0 & -\lambda & 1 \end{pmatrix} \quad \text{and} \quad \Omega = \begin{pmatrix} \omega & 0 & . & . & . & 0 \\ 0 & 1 & 0 & & & . \\ 0 & 0 & 1 & & & . \\ & & & . & & \\ . & & & & . & 0 \\ 0 & . & . & . & 0 & 1 \end{pmatrix}$$

ARX(1), COLS

Setting $\theta = (\lambda, \omega, \sigma^2)'$, this approximate bias can be written as $E[\hat{\lambda} - \lambda] = B_\theta(T^{-1}) + o(T^{-1})$, and then the first-order COLS estimator for the ARX case is given by

$$\hat{\lambda}_{COLS1} = \hat{\lambda} - B_{\hat{\theta}}(T^{-1}),$$

see Grubb and Symons (1987), Kiviet and Phillips (2003) and Kiviet and Phillips (2010).

- When calculating the COLS estimators for the ARX(1) model, an estimate of σ^2 is required as well as the coefficients.
- In the simulation results below, the OLS estimates are used for this - Kiviet and Phillips (2003) show that the COLS1 estimator is unbiased to order $O(T^{-1})$ under normal error terms using these estimates.

ARX(1), higher-order moment approximation

$$
\begin{aligned}
E[\hat{\lambda} - \lambda] &= B_\theta(T^{-2}) + o(T^{-2}) \\
B_\theta(T^{-2}) &= \sigma_u^2\{-q_{11} tr(Q\bar{Z}' C\bar{Z}) - q_1'\bar{Z}' C\bar{Z} q_1\} + \sigma_u^4\{-2q_{11}^2 tr(GG'C) \\
&\quad + 2q_{11}^2 tr(Q\bar{Z}' GG' C\bar{Z}) + 2q_{11}^2 tr(Q\bar{Z}' GG' C'\bar{Z}) \\
&\quad - 2q_{11}^2 tr(Q\bar{Z}' GG'\bar{Z} Q\bar{Z}' C\bar{Z}) - q_{11}^2 tr(Q\bar{Z}' C\bar{Z}) tr(Q\bar{Z}' GG'\bar{Z}) \\
&\quad + 6q_{11}(q_1'\bar{Z}' GG'(C + C')\bar{Z} q_1) - 6q_{11}(q_1'\bar{Z}' GG'\bar{Z} Q\bar{Z}' C\bar{Z} q_1) \\
&\quad - 6q_{11}(q_1'\bar{Z}' GG'\bar{Z} Q\bar{Z}' C'\bar{Z} q_1) - 3q_{11}(q_1'\bar{Z}' GG'\bar{Z} q_1) tr(Q\bar{Z}' C\bar{Z}) \\
&\quad - 3q_{11}(q_1'\bar{Z}' C\bar{Z} q_1) tr(Q\bar{Z}' GG'\bar{Z}) - 3(q_1'\bar{Z}' C\bar{Z} q_1)(q_1'\bar{Z}' GG'\bar{Z} q_1)\} \\
&\quad + \sigma_u^6\{8q_{11}^3 tr(GG'GG'C) - 2q_{11}^3 tr(GG'GG') tr(Q\bar{Z}' C\bar{Z}) \\
&\quad - 4q_{11}^3 tr(GG'C) tr(Q\bar{Z}' GG'\bar{Z}) - 20q_{11}^2(q_1'\bar{Z}' GG'\bar{Z} q_1) tr(GG'C) \\
&\quad - 10q_{11}^2(q_1'\bar{Z}' C\bar{Z} q_1) tr(GG'GG')\} \\
&\quad + \sigma_u^8\{-12q_{11}^4 tr(GG'C) tr(GG'GG')\} + o(T^{-2}),
\end{aligned}
$$

ARX(1), higher-order COLS

The bias approximation is given in Kiviet and Phillips (2010)

$$\hat{\lambda}_{COLS2} = \hat{\lambda} - B_{\hat{\theta}}(T^{-2}).$$

- The $\hat{\lambda}_{COLS2}$ estimator in (18) is unlikely to be unbiased to order $O(T^{-2})$ even with normal error terms.
- The issue is more difficult to resolve here than it was for the AR(1) case

Effect of skewness and kurtosis

- We can expect certain patterns in the behaviour of the simulated bias in the AR(1) and ARX(1) cases under non-Normality from the work of Bao and Ullah (2007) and Bao (2007)
- The bias in OLS estimation of the simple AR(1) does not depend on the skewness or kurtosis of error distributions to order $O(T^{-1})$
- This implies that the bias correction from COLS and QJ estimation will also not depend on skewness or kurtosis to this order.
- The same is not so for the ARX(1) - Bao and Ullah (2007)

Simulation setup

- All simulations were performed in Ox with 100,000 replications, using the simula class, and using the bootstrap and simulation classes due to Davidson (2005).
- The intercept was set as $\beta = 0$ in all cases
- In generating the fixed autoregressive exogenous variable x_t, the coefficient on x_{t-1} was chosen to be $\rho = 0.9$ in order to generate a relatively smooth series and to reflect the persistence of many economic time-series
- Within the ARX(1) model, the coefficient δ on the exogenous variable was chosen in a way that kept the signal-noise ratio constant across the parameters considered, as in Kiviet and Phillips (2003).

Simulation setup (continued)

- A signal-noise ratio of 19 was chosen, again as in Kiviet and Phillips (2003), implying a population R^2 of 0.95
- This implied that δ had to be

$$\delta = \left| \sqrt{[19(1-\lambda^2)-\lambda^2](1-\rho^2)\frac{1-\lambda\rho}{1+\lambda\rho}} \right| .$$

- So $|\lambda| \leq 0.975$

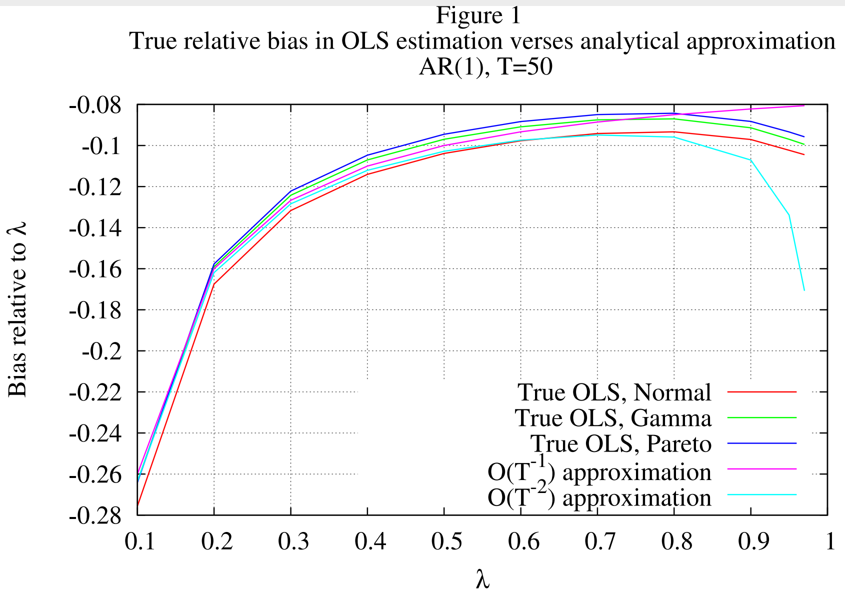


Figure 1
True relative bias in OLS estimation verses analytical approximation
AR(1), T=50
-0.08
-0.1
-0.12
-0.14
-0.16
-0.18
-0.2
-0.22
-0.24
-0.26
-0.28
Bias relative to λ
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
λ
True OLS, Normal
True OLS, Gamma
True OLS, Pareto
O(T^-1) approximation
O(T^-2) approximation

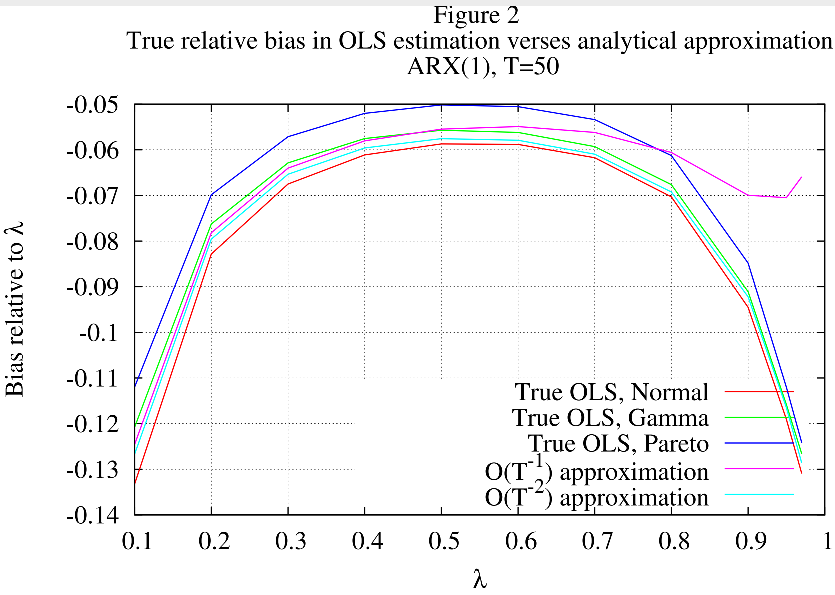
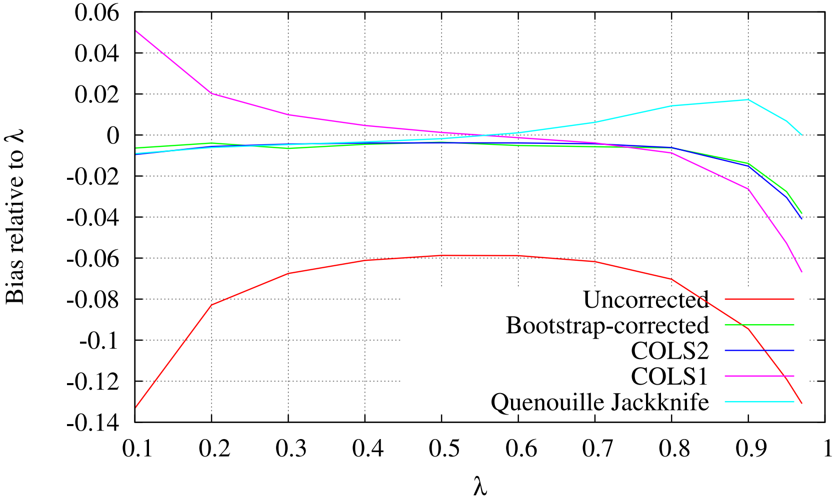


Figure 2
True relative bias in OLS estimation verses analytical approximation
ARX(1), T=50
Bias relative to λ
λ
-0.05
-0.06
-0.07
-0.08
-0.09
-0.1
-0.11
-0.12
-0.13
-0.14
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
True OLS, Normal
True OLS, Gamma
True OLS, Pareto
O(T-1) approximation
O(T-2) approximation

Figure 3
Relative bias comparisons
ARX, T=50, Normal disturbances



Bias reduction

- All estimators mentioned are preferable to ordinary least squares in terms of bias
- The Quenouille jackknife, residual bootstrap and higher-order COLS, $\hat{\lambda}_{COLS2}$, seem to do best overall.
- The estimators all seem fairly robust to non-Normal disturbances in terms of bias reduction, even COLS2
- The Quenouille jackknife seems to do very well near to the unit root.

Figure 4
Relative absolute bias comparisons (omitting OLS)
ARX, T=50, Normal disturbances

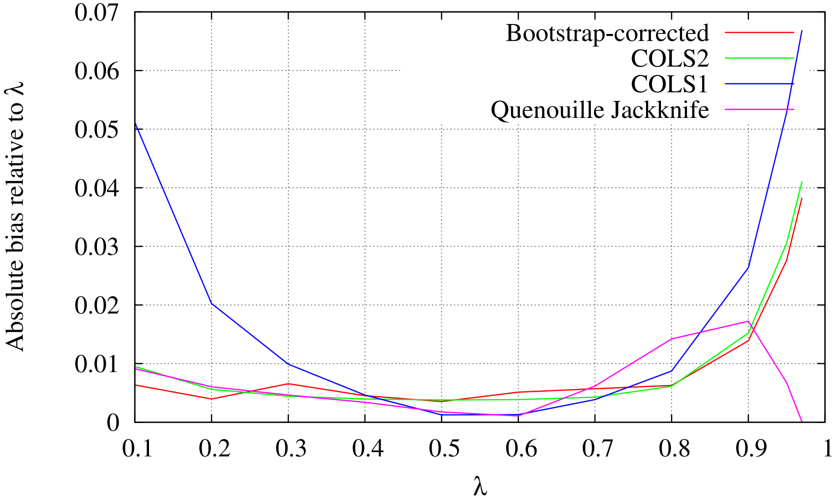


Figure 5
Relative absolute bias comparisons (omitting OLS)
ARX, T=50, Gamma disturbances

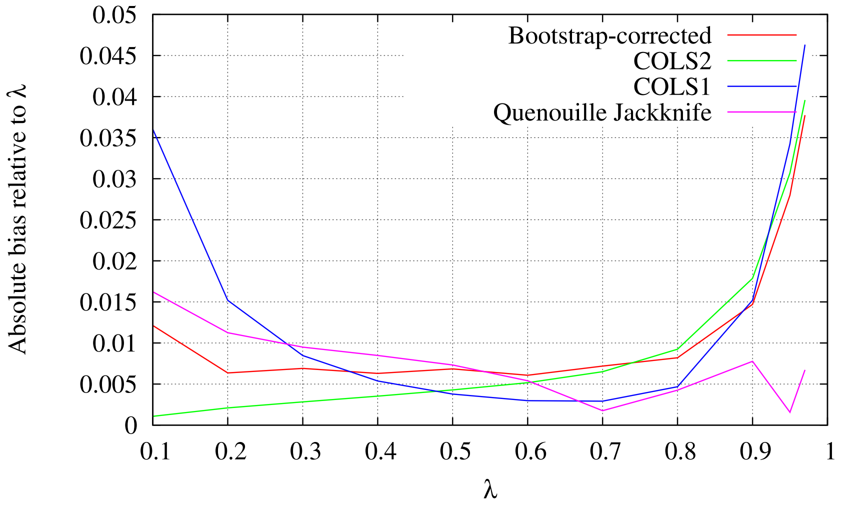
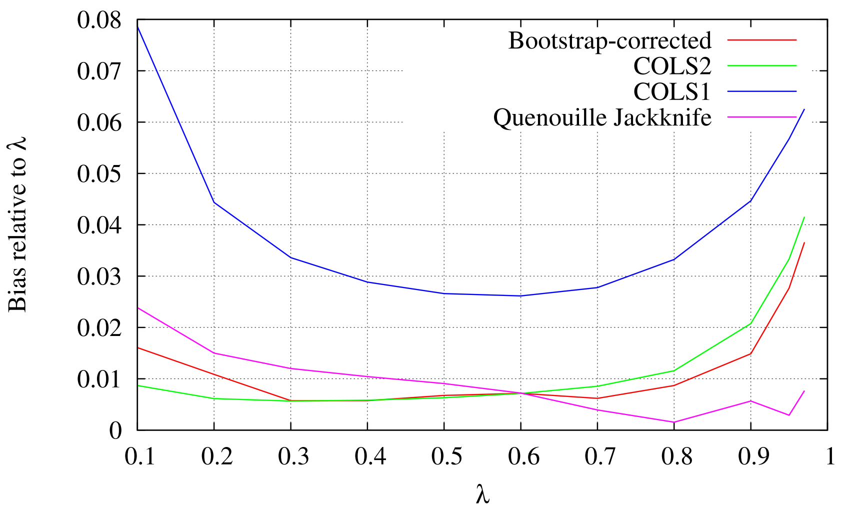


Figure 6
Relative bias comparisons (omitting OLS)
ARX, T=50, Pareto disturbances



MSE

- MSE of the COLS1 estimator $\hat{\lambda}_{COLS1}$ is typically the lowest
- though not in the Pareto case
- The next lowest MSE figures are for the bootstrap and the higher-order COLS estimator.
- The Quenouille jackknife is relatively inefficient.

Figure 7
MSE comparisons
ARX, T=50, normal disturbances

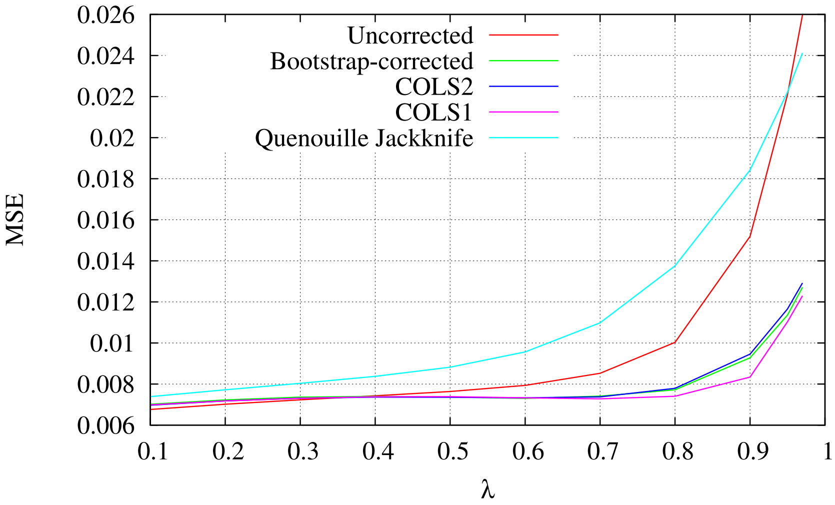
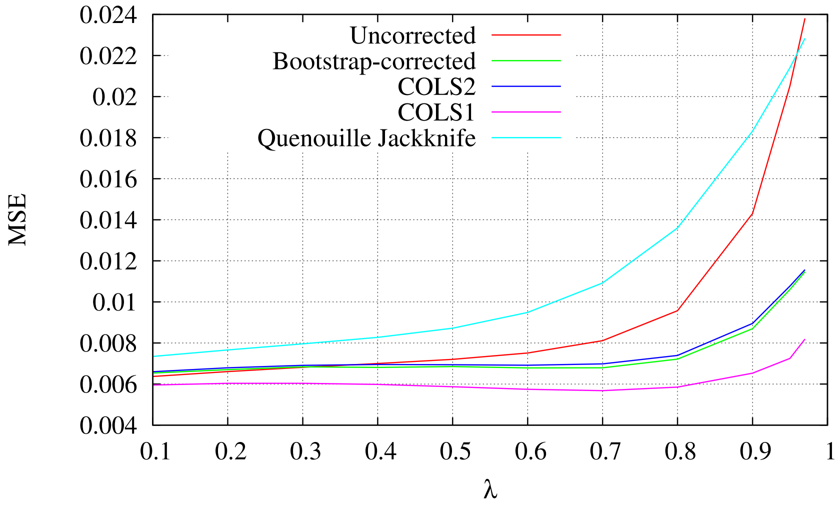


Figure 8
MSE comparisons
ARX, T=50, Gamma disturbances



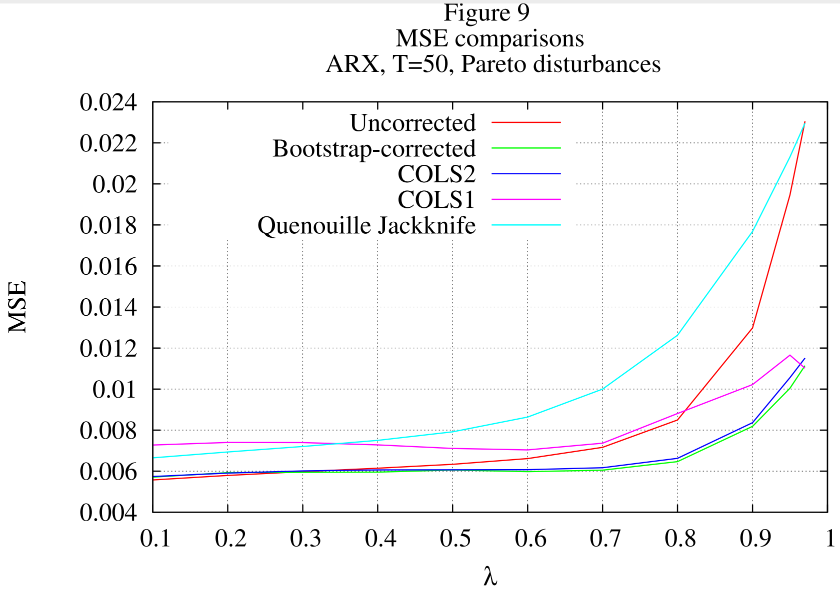


Figure 9
MSE comparisons
ARX, T=50, Pareto disturbances
Uncorrected
Bootstrap-corrected
COLS2
COLS1
Quenouille Jackknife
MSE
0.024
0.022
0.02
0.018
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
λ

Bias reduction

- COLS2 estimators $\hat{\lambda}_{COLS2}$ and $\hat{\lambda}_{COLS2,\star}$, and the alternative COLS estimator $\tilde{\lambda}_{COLS1}$, typically do best in terms of bias-correction
- followed by the Quenouille jackknife,
- then followed by the first-order COLS estimator and the bootstrap.
- From the analysis in Kiviet and Phillips (2003), one would expect $\tilde{\lambda}_{COLS1}$ to do better than $\hat{\lambda}_{COLS1}$ in terms of bias correction and worse in terms of MSE.
- The results for bias in the AR(1) are mixed.

Table 1: AR(1), mean, Normal, $T = 50$

λ	$\hat{\lambda}$	$\hat{\lambda}_{COLS1}$	$\tilde{\lambda}_{COLS1}$	$\hat{\lambda}_{COLS2}$	$\hat{\lambda}_{COLS2,\star}$	$\hat{\lambda}_Q$	$\hat{\lambda}_b$
0.1	0.0724	0.0968	0.0983	0.0972	0.0987	0.0984	0.0988
0.4	0.354	0.396	0.398	0.396	0.399	0.398	0.397
0.7	0.634	0.692	0.696	0.696	0.699	0.699	0.695
0.9	0.813	0.881	0.886	0.896	0.900	0.896	0.889
0.97	0.869	0.941	0.945	0.965	0.970	0.957	0.943

Table 2: AR(1), mean, Gamma, $T = 50$

λ	$\hat{\lambda}$	$\hat{\lambda}_{COLS1}$	$\tilde{\lambda}_{COLS1}$	$\hat{\lambda}_{COLS2}$	$\hat{\lambda}_{COLS2,\star}$	$\hat{\lambda}_Q$	$\hat{\lambda}_b$
0.1	0.0736	0.0981	0.0996	0.0985	0.100	0.0983	0.0985
0.4	0.357	0.399	0.401	0.399	0.402	0.396	0.397
0.7	0.639	0.697	0.701	0.701	0.704	0.696	0.697
0.9	0.818	0.887	0.891	0.902	0.906	0.893	0.889
0.97	0.874	0.946	0.951	0.982	0.986	0.953	0.944

Table 3: AR(1), mean, Pareto, $T = 50$

λ	$\hat{\lambda}$	$\hat{\lambda}_{COLS1}$	$\tilde{\lambda}_{COLS1}$	$\hat{\lambda}_{COLS2}$	$\hat{\lambda}_{COLS2,\star}$	$\hat{\lambda}_{Q}$	$\hat{\lambda}_{b}$
0.1	0.0736	0.0980	0.0996	0.0984	0.0999	0.0987	0.0991
0.4	0.358	0.400	0.402	0.400	0.403	0.398	0.399
0.7	0.641	0.699	0.703	0.703	0.706	0.698	0.698
0.9	0.821	0.890	0.894	0.900	0.905	0.895	0.890
0.97	0.877	0.950	0.954	0.999	1.003	0.956	0.945

MSE

- The results support a theoretical finding in Kiviet and Phillips (2003): when $-0.811 < \lambda < 0.411$ and errors are Normal, the MSE of $\hat{\lambda}_{COLS1}$ should be greater than that of $\hat{\lambda}$.
- The MSE of the higher-order COLS estimator is very large for values of λ approaching the unit root.
- The MSE for the Quenouille jackknife is also high.
- MSE of the COLS1 estimator $\hat{\lambda}_{COLS1}$ is typically the lowest.

Table 4: AR(1), MSE, Normal, $T = 50$

λ	$\hat{\lambda}$	$\hat{\lambda}_{COLS1}$	$\tilde{\lambda}_{COLS1}$	$\hat{\lambda}_{COLS2}$	$\hat{\lambda}_{COLS2,\star}$	$\hat{\lambda}_Q$	$\hat{\lambda}_b$
0.1	0.0203	0.0219	0.0221	0.0219	0.0221	0.0236	0.0222
0.4	0.0198	0.0199	0.0201	0.0201	0.0202	0.0228	0.0203
0.7	0.0177	0.0151	0.0152	0.0156	0.0157	0.0203	0.0155
0.9	0.0170	0.0109	0.0108	0.0476	0.0477	0.0186	0.0114
0.97	0.0179	0.00940	0.00922	2.36	2.36	0.0176	0.00974

Table 5: AR(1), MSE, Gamma, $T = 50$

λ	$\hat{\lambda}$	$\hat{\lambda}_{COLS1}$	$\tilde{\lambda}_{COLS1}$	$\hat{\lambda}_{COLS2}$	$\hat{\lambda}_{COLS2,\star}$	$\hat{\lambda}_Q$	$\hat{\lambda}_b$
0.1	0.0174	0.0187	0.0189	0.0188	0.0189	0.0245	0.0186
0.4	0.0161	0.0160	0.0161	0.0161	0.0163	0.0217	0.0161
0.7	0.0144	0.0120	0.0121	0.0125	0.0126	0.0180	0.0123
0.9	0.0146	0.00897	0.00894	0.0450	0.0451	0.0162	0.00951
0.97	0.0160	0.00807	0.00792	3.32	3.32	0.0161	0.00838

Table 6: AR(1), MSE, Pareto, $T = 50$

λ	$\hat{\lambda}$	$\hat{\lambda}_{COLS1}$	$\tilde{\lambda}_{COLS1}$	$\hat{\lambda}_{COLS2}$	$\hat{\lambda}_{COLS2,\star}$	$\hat{\lambda}_{Q}$	$\hat{\lambda}_{b}$
0.1	0.0165	0.0178	0.0179	0.0548	0.0179	0.0243	0.0179
0.4	0.0153	0.0152	0.0154	0.0676	0.0155	0.0216	0.0153
0.7	0.0137	0.0114	0.0115	0.0134	0.0121	0.0178	0.0116
0.9	0.0137	0.00837	0.00836	0.0474	1.76	0.0159	0.00878
0.97	0.0150	0.00753	0.00741	0.0801	137	0.0159	0.00792

Bias reduction (1)

- We fix λ_1 at 0.5 and consider $\lambda_2 \in \{0.05, 0.1, 0.2, 0.3, 0.4\}$ to explore the effect of moving nearer to the nonstationary region.
- The uncorrected OLS estimators $\hat{\lambda}_1$ and $\hat{\lambda}_2$ do worse as we move towards the nonstationary region.
- The bias can be substantial. In Table 7 the bias in OLS estimation of λ_2 is around 20% in the case nearest to nonstationarity.
- This suggests that the problem of bias near to the nonstationary region is greater in the higher-order model than in the AR(1).

Bias reduction (2)

- The bootstrap and Quenouille jackknife tend to do best in terms of bias-correction
- For Gamma and Pareto disturbances the COLS1 estimator does best in estimation of λ_2 near the nonstationary region
- For Pareto disturbances the COLS1 estimator is always best in estimating λ_1.

Table 7: AR(2), mean, Normal, $T = 50$, $\lambda_1 = 0.5$

λ_2	$\hat{\lambda}_1$	$\hat{\lambda}_{1,COLS}$	$\hat{\lambda}_{1,Q}$	$\hat{\lambda}_{1,b}$
0.05	0.465	0.494	0.503	0.498
0.1	0.463	0.493	0.503	0.497
0.2	0.459	0.491	0.505	0.497
0.3	0.454	0.487	0.506	0.495
0.4	0.444	0.479	0.505	0.488

$\hat{\lambda}_2$	$\hat{\lambda}_{2,COLS}$	$\hat{\lambda}_{2,Q}$	$\hat{\lambda}_{2,b}$
0.00452	0.0449	0.0497	0.0460
0.0502	0.0942	0.0993	0.0958
0.141	0.192	0.198	0.194
0.231	0.289	0.295	0.291
0.319	0.384	0.387	0.385

Table 8: AR(2), mean, Gamma, $T = 50$, $\lambda_1 = 0.5$

λ_2	$\hat{\lambda}_1$	$\hat{\lambda}_{1,COLS}$	$\hat{\lambda}_{1,Q}$	$\hat{\lambda}_{1,b}$
0.05	0.468	0.497	0.502	0.499
0.1	0.466	0.497	0.503	0.498
0.2	0.463	0.495	0.504	0.497
0.3	0.457	0.491	0.506	0.495
0.4	0.447	0.482	0.504	0.489

$\hat{\lambda}_2$	$\hat{\lambda}_{2,COLS}$	$\hat{\lambda}_{2,Q}$	$\hat{\lambda}_{2,b}$
0.00663	0.0472	0.0463	0.0479
0.0526	0.0969	0.0957	0.0972
0.144	0.196	0.194	0.196
0.234	0.293	0.291	0.293
0.323	0.389	0.383	0.386

Table 9: AR(2), mean, Pareto, $T = 50$, $\lambda_1 = 0.5$

λ_2	$\hat{\lambda}_1$	$\hat{\lambda}_{1,COLS}$	$\hat{\lambda}_{1,Q}$	$\hat{\lambda}_{1,b}$
0.05	0.468	0.498	0.502	0.499
0.1	0.467	0.497	0.503	0.499
0.2	0.463	0.495	0.505	0.499
0.3	0.458	0.491	0.506	0.494
0.4	0.448	0.483	0.505	0.489

$\hat{\lambda}_2$	$\hat{\lambda}_{2,COLS}$	$\hat{\lambda}_{2,Q}$	$\hat{\lambda}_{2,b}$
0.00913	0.0499	0.0486	0.0479
0.0553	0.0997	0.0980	0.0974
0.147	0.199	0.196	0.196
0.238	0.297	0.293	0.294
0.327	0.393	0.386	0.387

MSE

- The Quenouille jackknife is still relatively inefficient, but it does not seem as bad as in the AR(1) and ARX(1) cases.
- The COLS estimator seems to dominate in terms of MSE.
- It typically has the lowest MSE
- MSE of the COLS1 estimator $\hat{\lambda}_{COLS1}$ is typically the lowest

Table 10: AR(2), MSE, Pareto, $T = 50$, $\lambda_1 = 0.5$

λ_2	$\hat{\lambda}_1$	$\hat{\lambda}_{1,COLS}$	$\hat{\lambda}_{1,Q}$	$\hat{\lambda}_{1,b}$	$\hat{\lambda}_2$	$\hat{\lambda}_{2,COLS}$	$\hat{\lambda}_{2,Q}$	$\hat{\lambda}_{2,b}$
0.05	0.0198	0.0192	0.0280	0.0241	0.0197	0.0211	0.0287	0.0238
0.1	0.0202	0.0195	0.0286	0.0215	0.0201	0.0211	0.0289	0.0216
0.2	0.0209	0.0199	0.0297	0.0202	0.0208	0.0210	0.0293	0.0209
0.3	0.0215	0.0202	0.0309	0.0206	0.0216	0.0207	0.0297	0.0210
0.4	0.0224	0.0203	0.0320	0.0225	0.0226	0.0201	0.0303	0.0268

- We can compare estimates of $\sum_{i=1}^{2} \lambda_i$ using Tables 7-9
- The Quenouille jackknife does best in terms of bias reduction, but with high MSE
- The bootstrap seems to do best overall
- It does better in all but one case under Normality, and in all cases under Gamma disturbances.
- Under Pareto disturbances the COLS1 estimator does best in three out of five cases, including the case that is nearest to nonstationarity.

To summarise bias robustness we calculate the following:

$$\text{Bias Robustness} = (\sum_{\lambda \in \Lambda} z_\lambda^p)^{\frac{1}{p}},$$

where $\Lambda = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.97\}$, and where z_λ is the relative bias in estimation of λ. The MSE robustness is summarised in a similar way:

$$\text{MSE Robustness} = (\sum_{\lambda \in \Lambda} Z_\lambda^p)^{\frac{1}{p}},$$

where Z_λ is the MSE in estimation of λ divided by the MSE in OLS estimation of λ.

ARX(1)

- For Normal and Gamma errors in the ARX(1) model, the bootstrap is prefered over the COLS1 and COLS2 estimators in terms of bias robustness,
- For Normal and Gamma errors in the ARX(1) model, COLS estimators do better in terms MSE robustness.
- For Pareto errors in the ARX(1) model, the bootstrap does better than COLS1 and marginally better than COLS2 in terms of both bias and MSE robustness (see earlier figures)

Table 11: ARX(1), Bias Robustness, Normal, $T = 50$

p	$\hat{\lambda}$	$\hat{\lambda}_{COLS1}$	$\hat{\lambda}_{COLS2}$	$\hat{\lambda}_{Q}$	$\hat{\lambda}_{b}$
2	0.273	0.0917	0.0465	0.0263	0.0435
10	0.142	0.0673	0.0411	0.0175	0.0383
100	0.133	0.0669	0.0411	0.0172	0.0383

Table 12: ARX(1), MSE Robustness, Normal, $T = 50$

p	$\hat{\lambda}$	$\hat{\lambda}_{COLS1}$	$\hat{\lambda}_{COLS2}$	$\hat{\lambda}_{Q}$	$\hat{\lambda}_{b}$
2	3.162	2.777	2.807	3.684	2.812
10	1.259	1.199	1.198	1.519	1.204
100	1.023	1.037	1.036	1.371	1.043

AR(1)

- For Normal errors in the AR(1) model, the bootstrap does better than COLS1 in terms of bias robustness, but COLS1 does better in terms of MSE robustness.
- For Gamma and Pareto errors, COLS1 performs better in both the bias and MSE robustness measures (exception: MSE robustness with gamma errors and $\rho = 100$).
- The COLS2$\star$ estimator does very well in terms of bias robustness.

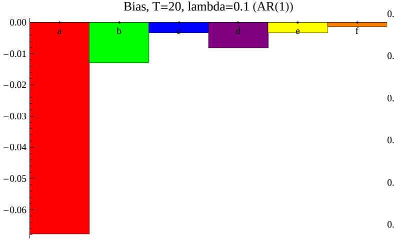
Table 13: AR(1), Bias Robustness, Normal, $T = 50$

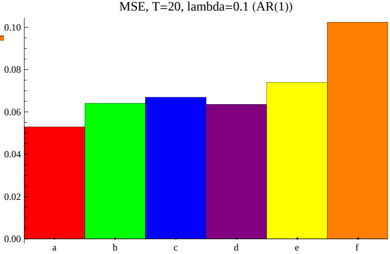
p	$\hat{\lambda}$	$\hat{\lambda}_{COLS1}$	$\tilde{\lambda}_{COLS1}$	$\hat{\lambda}_{COLS2}$	$\hat{\lambda}_{COLS2,\star}$	$\hat{\lambda}_{Q}$	$\hat{\lambda}_{b}$
2	0.439	0.0591	0.0378	0.0375	0.0155	0.0255	0.0382
10	0.276	0.0336	0.0254	0.0279	0.0132	0.0166	0.0282
100	0.276	0.0322	0.0254	0.0279	0.0132	0.0164	0.0282

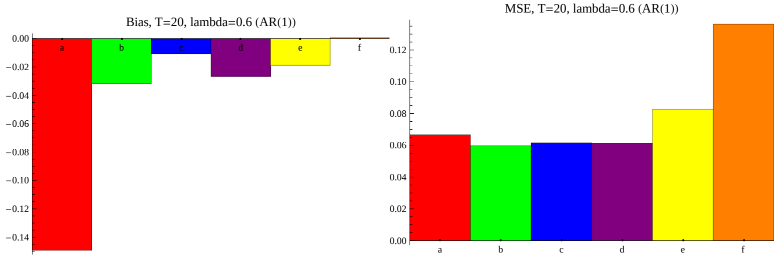
Table 14: AR(1), MSE Robustness, Normal, $T = 50$

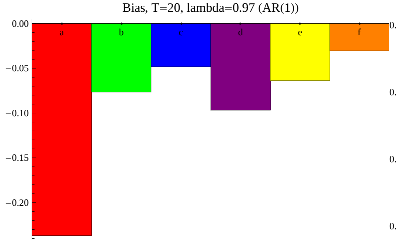
p	$\hat{\lambda}$	$\hat{\lambda}_{COLS1}$	$\tilde{\lambda}_{COLS1}$	$\hat{\lambda}_{Q}$	$\hat{\lambda}_{b}$
2	3.162	2.856	2.869	3.573	2.915
10	1.259	1.231	1.239	1.431	1.248
100	1.023	1.084	1.091	1.179	1.096

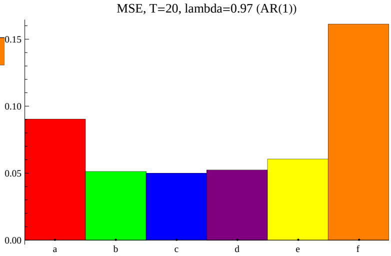
(a) OLS (b) COLS1 (c) Alternative COLS1 (d)Bootstrap (e) Quenouille Jackknife (f) Higher-order Jackknife






Bias, T=20, lambda=0.6 (AR(1))
0.00
−0.02
−0.04
−0.06
−0.08
−0.10
−0.12
−0.14
a
b
c
d
e
f
MSE, T=20, lambda=0.6 (AR(1))
0.12
0.10
0.08
0.06
0.04
0.02
0.00
a
b
c
d
e
f


Bias, T=20, lambda=0.97 (AR(1))
0.00
−0.05
−0.10
−0.15
−0.20
a
b
c
d
e
f


MSE, T=20, lambda=0.97 (AR(1))
0.15
0.10
0.05
0.00
a
b
c
d
e
f

Conculsions (1)

- The bootstrap generally does well
- Analytically corrected estimation based on $O(T^{-1})$ bias expansion is typically the most efficient method of bias reduction
- Significant improvements in bias reduction when moving from a COLS procedure based on mean approximation of order $O(T^{-1})$ to one based on approximation to order $O(T^{-2})$.
- The alternative higher-order jackknife, followed by the Quenouille jackknife, seem the best overall at bias reduction, but they are also the least efficient.

AR(1) + GARCH(1,1)

- Two GARCH(1,1) specifications, chosen to be as typical as possible.
- The AR(1)+GARCH(1,1) is frequently used to model log returns empirically.
- The values in Model ABD are the estimated coefficients in Angelidas et al. (2004) for daily log-returns on the FTSE 100 index, July 9th 1987 to October 18th 2002
- The values in Model BK, due to Brummelhuis and Kaufman (2007), were originally chosen by the authors as "typical values for GARCH models fitted to foreign exchange rates, stock indices, single stocks and 10-year government bonds".

Model ABD: $\bar{\omega} = 0.00000322$, $\alpha_1 = 0.116497$, $\alpha_2 = 0.859395$
Model BK: $\bar{\omega} = 0.000003$, $\alpha_1 = 0.05$, $\alpha_2 = 0.92$

Table 15: AR(1)+GARCH(1,1), mean $T = 50$, Model ABD

λ	$\hat{\lambda}$	$\hat{\lambda}_{COLS1}$	$\tilde{\lambda}_{COLS1}$	$\hat{\lambda}_Q$	$\hat{\lambda}_b$	$\hat{\lambda}_b^w$
0.1	0.0719	0.0962	0.0977	0.0975	0.0978	0.0975
0.4	0.352	0.394	0.396	0.396	0.395	0.396
0.7	0.632	0.690	0.693	0.696	0.692	0.692
0.9	0.810	0.879	0.883	0.893	0.884	0.885
0.97	0.865	0.937	0.942	0.953	0.938	0.939

Table 16: AR(1)+GARCH(1,1), MSE, $T = 50$, Model ABD

λ	$\hat{\lambda}$	$\hat{\lambda}_{COLS1}$	$\tilde{\lambda}_{COLS1}$	$\hat{\lambda}_Q$	$\hat{\lambda}_b$	$\hat{\lambda}_b^w$
0.1	0.0230	0.0249	0.0251	0.0296	0.0250	0.0250
0.4	0.0222	0.0225	0.0226	0.0278	0.0227	0.0226
0.7	0.0195	0.0167	0.0168	0.0233	0.0173	0.0170
0.9	0.0182	0.0118	0.0117	0.0201	0.0124	0.0125
0.97	0.0192	0.0104	0.0102	0.0192	0.0108	0.0107

Table 17: AR(1)+GARCH(1,1), mean $T = 50$, Model BK

λ	$\hat{\lambda}$	$\hat{\lambda}_{COLS1}$	$\tilde{\lambda}_{COLS1}$	$\hat{\lambda}_Q$	$\hat{\lambda}_b$	$\hat{\lambda}_b^w$
0.1	0.0722	0.0966	0.0981	0.0981	0.0982	0.0980
0.4	0.354	0.395	0.398	0.397	0.397	0.396
0.7	0.633	0.691	0.695	0.698	0.695	0.693
0.9	0.812	0.881	0.885	0.895	0.887	0.886
0.97	0.868	0.940	0.945	0.956	0.941	0.940

Table 18: AR(1)+GARCH(1,1), MSE, $T = 50$, Model BK

λ	$\hat{\lambda}$	$\hat{\lambda}_{COLS1}$	$\tilde{\lambda}_{COLS1}$	$\hat{\lambda}_Q$	$\hat{\lambda}_b$	$\hat{\lambda}_b^w$
0.1	0.0212	0.0230	0.0231	0.0256	0.0232	0.0230
0.4	0.0206	0.0208	0.0209	0.0244	0.0211	0.0210
0.7	0.0183	0.0156	0.0157	0.0212	0.0161	0.0161
0.9	0.0174	0.0112	0.0111	0.0191	0.0117	0.0118
0.97	0.0183	0.00970	0.00951	0.0181	0.0101	0.0102

Conclusions (2)

- Similar results to earlier
- All methods correct the bias well.
  - Surprisingly, the residual bootstrap for i.i.d. errors seems do well at bias correction even with the GARCH errors
- COLS continues to be the most efficient in terms of MSE

Possible future work

- Seems worthwhile to investigate the bias in higher-order models or vector autoregressive (VAR) models.
  - Relative biases of around 20% near the nonstationary region in an AR(2) at sample size 50
  - In a seemingly innocuous case such as $\lambda_1 = 0.5$ and $\lambda_2 = 0.2$ there are relative biases of around 9% and 30% respectively.
- Theoretical results on analytically corrected estimation when errors are non-i.i.d. e.g. GARCH
  - Iglesias and Phillips (2008) examine QML estimation in the strictly stationary known-mean case where $y_t = \lambda y_{t-1} + u$ and where errors are Guassian ARCH/GARCH, in particular the $O(T^{-1})$ bias in estimation of λ is still $\frac{-2\lambda}{T}$ in the $GARCH(p, q)$ case.
  - Currently no similar results for autoregressive models that include a constant or added exogenous regressors.

Angelidas, T., Benos, A., Degiannakis, S., 2004. The use of GARCH models in VaR estimation. Statistical Methodology 1, 105–128.

Bao, Y., 2007. The approximate moments of the least squares estimator for the stationary autoregressive model under a general error distribution. Econometric Theory 23(05), 1013–1021.

Bao, Y., Ullah, A., 2007. The second-order bias and mean-squared error of estimators in time-series models. Journal of Econometrics 140(2), 650–669.

Brummelhuis, R., Kaufman, R., 2007. Time-scaling for GARCH(1,1) processes. The Journal of Risk 9(4), 39–94.

Davidson, J., 2005. An Ox class for implementing bootstrap tests, documentation for Bootstrap class.

Dhaene, G., Jochmans, K., Thuysbaert, B., 2006. Jackknife bias reduction for nonlinear dynamic panel data models with fixed effects, presented at the Econometric Society European Meeting 2006.

Godfrey, L., Tremayne, A., 2005. The wild bootstrap and heteroskedasticity-robust tests for serial correlation in dynamic regression models. Computational Statistics and Data Analysis 49(2), 377–395.

Goncalves, S., Killian, L., 2004. Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. Journal of Econometrics 123(1), 89–120.

Grubb, D., Symons, J., 1987. Bias in regressions with a lagged dependent variable. Econometric Theory 3, 371–386.

Iglesias, E., Phillips, G., 2008. Finite sample theory of QMLE in arch models with dynamics in the mean equation. Journal of Time Series Analysis 29(4), 719–737.

Kendall, M., 1954. Note on bias in the estimation of autocorrelation. Biometrika 61, 403–404.

Kiviet, J., Phillips, G., 2003. Improved coefficient and variance estimation in stable first-order dynamic regression models, UvA Econometrics Discussion Paper.

Kiviet, J., Phillips, G., 2010. Higher-order asymptotic expansions of the least-squares estimation bias in first-order dynamic regression models. Forthcoming at Computational Statistics and Data Analysis, 6th Special Issue on Computational Econometrics.

Marriott, F., Pope, J., 1954. Bias in the estimation of autocorrelations. Biometrika 61, 393–403.