

**Small-sample performance of linear estimators
in autoregressions with common seasonals:
An example for Bayesian Monte Carlo evaluation
of classical estimators**

Robert M. Kunst

University of Vienna

and

Institute for Advanced Studies Vienna

Prepared for the RSS 2006 Meeting, Belfast, September 10-14

Table of contents

1. The estimation problem
2. Monte Carlo simulations for assessing estimators
3. Introducing Bayesian MC simulations
4. Design of the MC experiments
5. Results

The estimation problem

Consider the model

$$y_t = m + Ax_t + By_{t-1} + u_t, \quad t = 1, \dots, T,$$

where y_t , y_{t-1} , m , u_t are n -vectors,

A is an $n \times 3$ -matrix,

B is an $n \times n$ -matrix,

x_t is a 3-vector of seasonal constants in trigonometric form.

This model may appear in panels of quarterly data.

A rank restriction on A

In general, the matrix A will have a rank of 3 for $n \geq 3$.
 $\text{rk}(A) < \min(n, 3)$ indicates that characteristics in the seasonal cycle are common across the individual dimension.

In particular, $\text{rk}(A) = 1$ describes the case that the seasonal cycle is identical across individuals.

Means ('effects') and dynamics (here expressed by the simple first-order autoregressive matrix B) may be individual-specific.

B is a diagonal matrix

If B is a general $n \times n$ -matrix, estimation of the model is fairly standard.

If u is Gaussian *iid*, the maximum-likelihood estimator is the least-squares equation-by-equation (that is, individual-by-individual) estimator.

The Kruskal theorem guarantees that the covariance structure of u plays no role.

Similarly, for rank-restricted A , standard reduced-rank regression (RRR) yields the maximum-likelihood estimator.

If B is diagonal, such that individuals depend on their own lags only, the conditions of the Kruskal theorem are violated.

Sequences of linear estimators

Sequences of linear and consistent estimators will not be efficient but they are consistent and useful in practice.

Such a sequence may start from RRR without restricting B . The estimate for A can be decomposed as $A = \alpha\beta'$, and least squares yields final estimates of α , m , B .

This estimator will be called EST2 in the following. Part of its inefficiency is due to the over-parameterized vector autoregression in its first step.

Alternatively, one may start by estimating b_j without restricting A . Continue with a RRR, and perform one more round of least squares to estimate α , m , B . This estimator will be called EST1 in the following.

The efficient brute-force backdrop

Still assuming that the errors are Gaussian and that their covariance matrix is scalar, one may also subject the model to non-linear maximization of the likelihood.

This estimator will be asymptotically efficient.

It is far more interesting how it will behave in smaller samples.

Likelihood estimation may also be time-consuming.

This estimator will be called **EST3** in the following.

Analytical results on the estimators are difficult to obtain

Derivatives of the likelihood would be needed for all parameters.

The literature on reduced-rank regression offers no results for the coefficients in B .

Even if one could derive the asymptotic properties of the methods **EST1** and **EST2**, it is doubtful that those would help for smaller samples.

It is preferable to rely on Monte Carlo simulations.

How are MC evaluations of estimators designed?

Statistical journals offer a multitude of role models for MC estimator evaluation.

In a meta-evaluation of around 200 articles (*JASA*, *Journal of Econometrics* etc.), the following patterns emerged:

- 1) Designs of simulation experiments rely on *one(!)* to hundreds of assumed parameter values.
- 2) Parameter values can be inspired from estimation based on 'realistic' or 'typical' data. They can also conform to a grid. One may also set them arbitrarily, conforming to an interest in exploring certain constellations.

How is the quality of an estimator assessed?

Criteria for evaluating the quality of estimators vary a lot.

The following can be found (among others):

- 1) means, variances, mean squared errors, standard deviations;
- 2) medians, interquartile ranges;
- 3) higher moments (rare);
- 4) empirical CDF plots.

Most statistics are shown in tables.

For grid designs (and CDF), they are also shown in graphs.

What follows the collection of MC evidence?

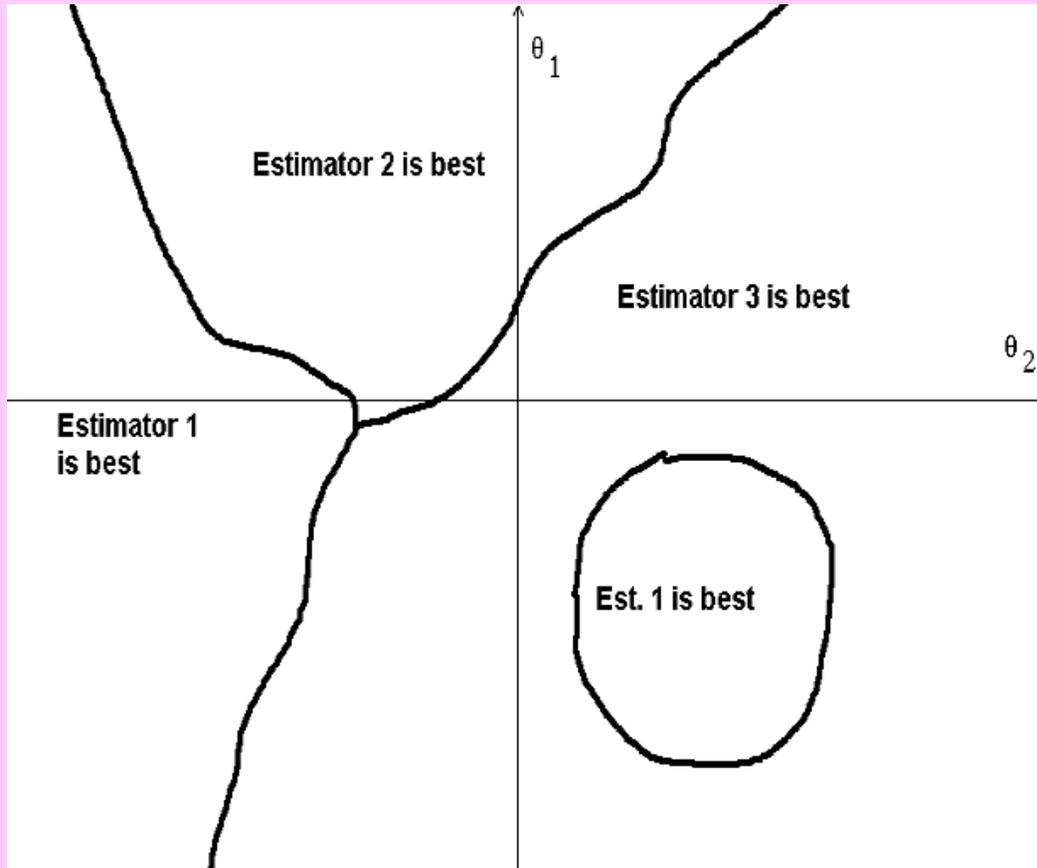
Most reported simulations do not yield a general dominance of one (out of two or many more) estimators.

Nevertheless, around $\frac{3}{4}$ of all articles conclude with some kind of *recommendation*.

In the remaining $\frac{1}{4}$ of articles, there is dominance or no such recommendation is provided.

Typically, the ranking of estimators varies across designs. An estimator may be good for most designs but extremely bad for some designs with special parameter values, such as ' $\rho < 0$ '.

A typical outcome



Graph assumes 2-dimensional parameter and exhaustive design.

Resolving conflict

- 1) More than half of the considered articles give a pre-test style recommendation (“If the autocorrelation of the errors is positive and small, estimator 1 is best”).
Problem: the implicitly recommended pre-testing defines a new estimator that was not evaluated.
- 2) Around 20% of the articles give a Bayes-style conclusion (“In most economic time series, autocorrelation tends to be positive”), thereby assigning more importance to certain regions of the parameter space.

A formalization of standard MC

The targeted quantity is the moment characteristic

$$g(\hat{\theta}, \theta, T) = \int_{A^T} g^* (\hat{\theta}(x) - \theta) f_{\theta}^T(x) dx,$$

where x stands for the sampled T -vectors in A^T .

The function g^* may be squares (or a multivariate version) for the variance, an absolute value for absolute first moment etc.

$g(\theta, \theta, T) = 0$, such that guessing the true value is perfect.

If an estimator yields a smaller g integral, it is 'better'.

Measures such as the absolute bias and standard error are similarly defined by transforms of integral (|.| and root).

(An analogous exposition can be presented for fractile-type characteristics)

Approximating the g integral

This can be conducted by drawing r replications from the density f_{θ}^T . Then, evaluate

$$r^{-1} \sum_{j=1}^r g * (\hat{\theta}^{(j)} - \theta)$$

and use convergence by the law of large numbers.

The approximate integral depends on the estimator, the true value θ , the sample size T , the number of replications, and it is random. The limit does not depend on r and ω , but on the first three arguments.

Bayesian MC evaluation

For a Bayesian evaluation of estimators, a probability density $h(\theta)$ is defined over $\theta \in \Theta$.

The aim becomes to minimize

$$\int_{\Theta} h(\theta) g(\hat{\theta}, \theta, T) d\theta = \tilde{g}(\hat{\theta}, T),$$

which allows a complete ranking of estimators. Then, the ranking depends on T (acceptable) and on h (problematic).

Dependence on h can be explored by sensitivity analysis.

Bayes estimation and Bayes MC

While both applications are similar in spirit, there are some distinctive features.

Conjugate priors may be important for Bayes estimation but they are irrelevant in MC.

Improper priors may be acceptable for Bayes estimation but they cannot be used in MC.

Priors h should be guided by what may be a likely value for the parameter in a potential empirical application, not by what may be a likely value in the considered application, as we do not really have a data set in MC.

Designs of experiments

$r=10,000$ replications are generated for T -samples of

$$y_t = m + Ax_t + By_{t-1} + u_t, \quad t = 1, \dots, T.$$

B is diagonal. Its elements are independent $U(0,1)$ draws. Some experiments were run with $U(-1,1)$ but negative autocorrelation sometimes causes extreme deterioration of *all* competing estimators, as it interferes with the negative autocorrelation in x .

The errors u are simply *iid* $N(0,1)$.

The reduced-rank matrix A

The $n \times 3$ -matrix A has rank one.
Therefore, it can be represented as

$$A = \alpha\beta',$$

with an n -vector α and a 3-vector β .
All elements are drawn from $N(1,1)$, except for the
first element in β , which is set at 1.
This standardizes the matrix A and avoids most
near-rank-zero replications.

T, r, g, n

T is taken from the set $\{100, 200, 500, 1000\}$.

r is reduced to 1000 for the nonlinear likelihood estimator EST3, due to excessive computer time.

Regarding g , means, standard errors, medians, upper and lower quartiles are evaluated.

Ultimately, there will be an emphasis on interquartile ranges.

Some replications cause very bad estimates, and robust measures of precision are required.

Dimensions $n=2,3,4,10$ are explored.

An alternative heteroskedastic design

As an instrument of sensitivity analysis, we also consider a heteroskedastic h design.

Individual error variances follow half-normal distributions, which yields quite extreme heteroskedasticity.

It would be easy to construct other estimators for this case, t.ex. by introducing another stage in EST1/2, rotating the variables in the style of SUR.

Or use a non-linear estimator with general diagonal variance matrix.

Such extensions are not in focus here, as we wish to compare EST1, EST2, EST3.

	$T=100$			$T=200$		
	EST1	EST2	EST3	EST1	EST2	EST3
m_1	0.2312	0.2382	0.2265	0.1617	0.1658	0.1585
m_2	0.2361	0.2404	0.2221	0.1606	0.1659	0.1662
α_1	0.1367	0.1398	0.1408	0.0950	0.0983	0.0978
α_2	0.1358	0.1410	0.1344	0.0931	0.0972	0.0925
β_2	0.1826	0.1912	0.1806	0.1253	0.1324	0.1256
β_3	0.2035	0.2225	0.2001	0.1450	0.1573	0.1463
b_1	0.1007	0.1037	0.0922	0.0674	0.0712	0.0657
b_2	0.0979	0.1012	0.0931	0.0665	0.0710	0.0647

Interquartile ranges. $n=2$. Best: green, 2nd best: blue.

	$T=100$			$T=200$		
	EST1	EST2	EST3	EST1	EST2	EST3
m_1	-0.0001	-0.0002	-0.0004	0.0007	0.0018	-0.0001
m_2	0.0005	-0.0006	-0.0085	-0.0002	-0.0005	-0.0060
α_1	0.0079	0.0094	0.0076	0.0046	0.0059	0.0076
α_2	0.0095	0.0115	0.0118	0.0040	0.0047	0.0095
β_2	0.0098	0.0116	0.0020	0.0037	0.0039	0.0001
β_3	-0.0106	-0.0144	-0.0053	-0.0077	-0.0100	-0.0132
b_1	0.0147	0.0163	0.0157	0.0088	0.0101	0.0089
b_2	0.0153	0.0178	0.0150	0.0075	0.0089	0.0071

Medians. $n=2$. Best: green, 2nd best: blue.

General impression

For $T=100$, the nonlinear ML estimator EST3 is supported by most figures. For larger T , performance of EST1 and EST3 is comparable.

EST1 appears to be better than EST2.

Dominance is less pronounced for non-robust moment characteristics.

Non-convergence of the non-linear optimization in EST3 occurred in some replications. These replications were eliminated, which biases the outcome in favor of EST3.

EST3 is enormously time-consuming.

For $T=r=1000$, a standard PC needs 40 hours.

Back to scalar evaluation

In order to allow a condensed assessment of the three estimators, the following tables focus on a scalar measure g , which is defined as the 2-norm for vectors of **interquartile ranges** of parameter estimates.

This measure fulfills the requirement of robustness, while it fails to reflect bias.

Medians are considered for the bias problem.

T	EST1	EST2	EST3
100	0.4907	0.5109	0.4779
200	0.3396	0.3557	0.3414
500	0.2125	0.2197	0.2090
1000	0.1483	0.1532	0.1495

Summary evaluation of the homoskedastic experiment based on interquartile range. $n=2$.

T	EST1	EST2	EST3
100	0.4018	0.4278	0.4466
200	0.2782	0.2929	0.3154
500	0.1727	0.1811	0.1878
1000	0.1210	0.1263	0.1317

Summary evaluation of the heteroskedastic experiment based on interquartile range. $n=2$.

T	EST1	EST2
100	0.5265	0.5695
200	0.3659	0.3898
500	0.2264	0.2409
1000	0.1587	0.1676

Summary evaluation of the homoskedastic $n=3$ experiment based on interquartile range.

General impression

EST1 does better than EST2.

It is preferable not to start from an over-specified vector autoregression.

EST3 is time-consuming and often fails to converge. Even for convergent cases, it is approximately as accurate as EST1.

There is no obvious dependence of ranking on T .

Replacing the 2-norm by a weighted criterion may deserve consideration, as β may be more interesting than m .

Conclusions about Bayes MC

Bayes MC helps considerably in providing a clear and global ranking among estimators.

In principle, Bayes MC explores the performance of estimators for the whole parameter space.

Traditional MC or forced weighting of parameter space areas may help in gaining insight on local ranking for pre-test estimators.

Sensitivity experiments are helpful. In this case, the experiment confirms the original result.