Monte Carlo Simulation for Econometricians

Monte Carlo Simulation for Econometricians

Jan F. Kiviet

Nanyang Technological University
Singapore
jfkiviet@ntu.edu.sg

University of Amsterdam
The Netherlands
j.f.kiviet@uva.nl

the essence of knowledge

Boston – Delft

Foundations and Trends® in Econometrics

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is J. F. Kiviet, Monte Carlo Simulation for Econometricians, Foundations and Trends® in Econometrics, vol 5, nos 1–2, pp 1–181, 2011

ISBN: 978-1-60198-538-5

Foundations and Trends® in Econometrics

Volume 5 Issues 1–2, 2011

Editorial Scope

Foundations and Trends® in Econometrics will publish survey and tutorial articles in the following topics:

- Identification
- Model Choice and Specification Analysis
- Non-linear Regression Models
- Simultaneous Equation Models
- Estimation Frameworks
- Biased Estimation
- Computational Problems
- Microeconometrics
- Treatment Modeling
- Discrete Choice Modeling
- Models for Count Data
- Duration Models
- Limited Dependent Variables
- Panel Data
- Dynamic Specification
- Inference and Causality
- Continuous Time Stochastic Models
- Modeling Non-linear Time Series
- Unit Roots
- Cointegration
- Latent Variable Models
- Qualitative Response Models
- Hypothesis Testing
- Interactions-based Models
- Duration Models
- Financial Econometrics
- Measurement Error in Survey Data
- Productivity Measurement and Analysis
- Semiparametric and Nonparametric Estimation
- Bootstrap Methods
- Nonstationary Time Series
- Robust Estimation

Information for Librarians

Foundations and Trends® in Econometrics, 2011, Volume 5, 4 issues. ISSN paper version 1551-3076. ISSN online version 1551-3084. Also available as a combined paper and online subscription.

Foundations and Trends® in
Econometrics
Vol. 5, Nos. 1–2 (2011) 1–181

DOI: 10.1561/0800000011

Monte Carlo Simulation for Econometricians

Jan F. Kiviet[1,2]

[1] *Division of Economics, School of Humanities and Social Sciences, Nanyang Technological University, Singapore 637332, jfkiviet@ntu.edu.sg*
[2] *Department of Quantitative Economics, Amsterdam School of Economics, University of Amsterdam, 1018 XE Amsterdam, The Netherlands, j.f.kiviet@uva.nl*

Abstract

Many studies in econometric theory are supplemented by Monte Carlo simulation investigations. These illustrate the properties of alternative inference techniques when applied to samples drawn from mostly entirely synthetic data generating processes. They should provide information on how techniques, which may be sound asymptotically, perform in finite samples and then unveil the effects of model characteristics too complex to analyze analytically. Also the interpretation of applied studies should often benefit when supplemented by a dedicated simulation study, based on a design inspired by the postulated actual empirical data generating process, which would come close to bootstrapping. This review presents and illustrates the fundamentals of conceiving and executing such simulation studies, especially synthetic but also more dedicated, focussing on controlling their accuracy, increasing their efficiency, recognizing their limitations, presenting their results in a coherent and palatable way, and on the appropriate interpretation of their actual findings, especially when the simulation study is used to rank the qualities of alternative inference techniques.

Contents

Preface and Overview

Since many decades much of the research in econometric theory is supported or illustrated by Monte Carlo simulation studies. Often the design of such studies follows particular patterns that have become traditional. Performing Monte Carlo studies is usually not taught as such in graduate schools. As a novice one is simply expected to imitate and extend relevant earlier studies published in the recent literature. Many scholars seem to think that setting up a Monte Carlo study is basically too self-evident to bother much about; apparently, it can be done without requiring a manual, because that does not seem available. Therefore, we try to present and illustrate the fundamentals of executing such studies here, pointing to opportunities not often utilized in current practice, especially regarding designing their general setup, controlling their accuracy, recognizing their shortcomings, presenting their results in a coherent and palatable way, and with respect to an appropriate and unprejudiced interpretation of their actual findings.

Monte Carlo simulation (abbreviated as MCS from now on) produces from random experiments rather straight-forward statistical inference on the properties of often very complex statistical inference techniques. So, it has an intrinsic recursive nature, because it employs statistical methods to explore statistical methods. Here, we will focus in particular on exploring the properties of classic econometric inference techniques by simulation. The major issues concerning these techniques are concisely characterized in Appendix A. In practice, they are usually applied to observational (i.e., non-experimental) data, employing methods and making probabilistic assumptions in a situation of high uncertainty regarding the appropriate model specification. Hence, in this context MCS examines complex techniques of statistics by rather simple techniques of statistics, aiming to produce knowledge on how to handle non-experimental data by experimentation.

At first sight such an approach may seem to be built on very weak methodological grounds, if not just being impossible. Indeed, its inherent circularity and apparent incompatibilities may easily lead to confusion. Reasons for that being that concepts such as sample and its sample size, estimators and their precision, test statistics and their significance levels, confidence regions and their coverage probabilities, and so on, play a role at two different levels, namely that of the econometric technique under investigation and that of the simulation inference produced on its properties. Therefore, we shall find it useful to develop a notation in which we carefully distinguish between the econometric issues under study and the statistical inference methods employed in MCS to interpret the simulation experiments. Such a distinction is nonstandard in the literature, but we think it is illuminating and certainly useful from a pedagogic point of view. For similar reasons we find it also highly instructive to use EViews programs for illustrations. Not because we appreciate the EViews programming language as such very much, but because it will prove to be most clarifying and convenient that in the computer sessions to follow we will as a rule have two data workfiles. One regarding samples to which the econometric techniques under study are applied, and one with usually a much larger sample size concerning the executed simulation experiments. To the latter we may immediately (or at a later stage, and step by step) apply any descriptive or inferential statistical techniques from the standard EViews menu deemed useful for interpreting the simulation findings.

In the first three sections the focus is on the basic tools of MCS, which are generating and transforming random numbers such as may arise in econometric analysis, and next assessing their moments, probability distributions and their quantiles numerically. We discuss and illustrate their use to produce MCS inference on the qualities of various specific econometric inference techniques, and to control the accuracy of the MCS results. Especially regarding the accuracy of MCS results we produce some findings that are seldom employed in practice. But also regarding some more familiar results we think that by carefully distinguishing in our notation between the statistical inference techniques under study and the statistical inference techniques employed to interpret the simulation experiments, we illuminate various aspects that are easily confused or overlooked otherwise. At the same time, by illustrating MCS to various of the standard tools of econometric inference, one may acquire a deeper and more tangible understanding of often rather abstract aspects of econometric theory. Not only does it illustrate the relevance and accuracy (and often the inaccuracy, thus limited relevance) of asymptotic theory and of the approximations it suggests. It will also help to sharpen the understanding of basic concepts such as bias and inconsistency, standard deviation and standard error, variance and mean squared error, the standard deviation of standard errors, nominal and actual significance levels, test size and power, and to appreciate how crucial (or occasionally trifling) the validity of particular standard assumptions (such as exogeneity, linearity, normality, independence, homoskedasticity) may be. Both producers and consumers of MCS results may appreciate the easy-to-use rules of thumb provided on the relationship between accuracy of the various obtained MCS results such as bias, median, RMSE, rejection probability, and the chosen number of replications or the Monte Carlo sample size.

After treating the basic tools of MCS, the focus of Section 4 is on the crucial elements of analyzing the properties of asymptotic test procedures by MCS. This involves verifying the extent of control over the type I error probability, establishing the test size, essential aspects of size correction when making power comparisons between competing test procedures. In Section 5 the focus is on various more general aspects of MCS, such as its history, possibilities to increase its efficiency and effectivity, whether synthetic random exogenous variables should be kept fixed over all the experiments or be treated as genuinely random and thus redrawn every replication. Here we also pay some attention to what we call a dedicated MCS study. Finally, it tries to provide a list of all methodological aspects that do affect MCS. We pay attention especially to those which are relevant when simulation results are used to rate various alternative econometric techniques. Most of these aspects receive very little attention in the majority of the currently published simulation studies. We list ten general methodological rules and aspirations, or rather commandments, to be followed when designing and executing Monte Carlo studies in econometrics when its purpose is an impartial validation of alternative inference techniques. Next we address the adverse effects sinning against these rules has.

The simulation techniques that we discuss in the first five sections are often addressed as naive or classic Monte Carlo methods. However, simulation can also be used not just for assessing the qualities of inference techniques, but also directly for obtaining inference in practice from empirical data. Various advanced inference techniques have been developed which incorporate simulation techniques. An early example of this is Monte Carlo testing, which corresponds to the (much later developed) parametric bootstrap technique. In the final Section 6 such techniques are highlighted, and a few examples of (semi-)parametric bootstrap techniques are given. This section also demonstrates that the bootstrap is not an alternative to MCS but just another practical — though usually asymptotic, and therefore probably inaccurate — inference technique, which uses simulation to produce econometric inference. If one wants to analyze the actual performance of bootstrap inference this can again be done by MCS, as we illustrate. Other advanced uses of simulation, such as in indirect inference or estimation by simulated moments methods or MCMC (Markov chain Monte Carlo) methods will not be covered here.

At the end of each section exercises are provided which allow the reader to immerse in performing and interpreting MCS studies. The material has been used extensively in courses for undergaduate and graduate students. The various sections contain illustrations which throw light on what uses can be made from MCS to discover the finite sample properties of a broad range of alternative econometric methods with a focus on the rather basic models and techniques. Just occasionally, we pay attention (by giving references) to how to condense the often rather extensive reporting on simulation findings by employing graphical 2D and 3D methods, which can even be extended to 4D by using animation. This, however, requires other software than provided by the EViews package.

Although Monte Carlo is practiced now for more than a century and started in fact long before computers were available by manually drawing repeatedly independent samples from a given population, there are no many texts that explain and thoroughly illustrate MCS and its foundations in detail for econometricians. In that respect we should however name at least the following relatively few exceptions. The relevant underlying theory for examining isolated inference techniques (estimators and test procedures) can be found in Hendry (1984, Section 16 of *Handbook of Econometrics*, Vol. II). Sections on Monte Carlo simulation can also be found in the econometrics textbooks by Davidson and MacKinnon (1993, Section 21), Hendry (1995, Section 3, Section 6), Hendry and Nielsen (2007, Section 16), and intermingled with bootstrap applications throughout most sections of Davidson and MacKinnon (2004). An initial study in Monte Carlo methodology, focussing on issues that are relevant when alternative inference methods are compared by Monte Carlo methods, is Kiviet (2007), which is extended here. A recent introduction to the basics of Monte Carlo methods, focussing in particular on random number generation, is Doornik (2006), published in a volume in which Davidson and MacKinnon (2006) provide an introduction for econometricians to the bootstrap. Further relatively recent introductions to the bootstrap are Horowitz (2003), Johnson (2001), and MacKinnon (2002, 2006). There are many more advanced bootstrap papers in the econometrics literature, see for instance Brown and Newey (2002). For other inference methods which involve simulation (i.e., where simulation is not just used to analyze the quality of inference but to help to produce inference), which are not covered here, such as Method of Simulated Moments, Indirect Inference, Markov Chain Monte Carlo, Gibbs Sampling, Simulated Annealing etc. see, for instance, Fishman (2006) and Gourieroux and Monfort (1996) and the overview in Greene (2012).

Acronyms and Symbols Used

AR(p)	autoregressive process of order p
ARX(p)	regression model with exogenous regressors X and lagged dependent variables up to order p
BSS	bootstrap simulation
CDF	cumulative distribution function
CLT	central limit theorem
DGP	data generating process
E	expectation
ECDF	empirical cumulative distribution function
EPDF	empirical probability distribution function
ϵ	relative precision
IQR	interquartile range
GLS	generalized least-squares
GMM	generalized method of moments
IID	identically and independently distributed
IV	instrumental variables
κ	kurtosis
LIE	law of iterated expectations

LLN	law of large numbers
λ	skewness
MCS	Monte Carlo simulation
MD	median
ML	maximum likelihood
MSE	mean squared error
NIID	normal and IID
NLS	nonlinear least-squares
OLS	ordinary least-squares
PDF	probability density function
PMC	pivotal Monte Carlo (test)
RMSE	root mean squared error
SD	standard deviation
SE	standard error
τ	absolute tolerance
TSLS	two-stage least-squares
UIID	uniform and IID
Var	variance
VSE	variance of the squared error

1

Introduction to Classic Monte Carlo Simulation

1.1 Main Purposes and Means

Computers have the facility to generate seemingly independent drawings from well-known discrete and continuous distributions, such as Bernoulli, binomial, uniform, normal, Student, etc. using a so-called pseudo random number generator. By transforming these random numbers it is also possible to obtain drawings from the distribution of complicated functions of such standard distributions, and thus for any econometric estimator of a parameter vector, and for its variance estimator and for related test statistics. The analytic assessment of the actual cumulative distribution functions, densities, quantiles or moments of these estimators and test statistics is usually intractable, because mostly they are highly nonlinear functions of the random disturbances, the often random regressor and instrumental variables and the model parameters. By using the computer to draw a large IID (identically and independently distributed) sample from such a complicated distribution, we can use this Monte Carlo sample to estimate its moments numerically, provided these exist, whereas a histogram of this sample establishes the empirical probability distribution. Likewise, the empirical counterparts of the cumulative distribution function (CDF), and if this exists the probability density function (PDF) can be assessed, and quantiles of the unknown distribution can be found by inverting the CDF. Of course, such Monte Carlo estimators of characteristics of an unknown distribution do entail estimation errors themselves. So, in order to be able to judge their inaccuracies, Monte Carlo results should — like all statistical inference — always be supplemented by appropriate corresponding standard errors, confidence regions, etc.

In this introductory section, we will not yet practice Monte Carlo simulation properly, but just illustrate the generation of random variables on a computer by EViews, and employ this to illustrate various basic aspects of the LLN (Law of Large Numbers) and the CLT (Central Limit Theorem), which jointly do not only form the backbone of asymptotic theory on econometric inference, but — as we shall soon find out — also of the interpretation and the control of the actual accuracy of MCS (Monte Carlo simulation) findings.

1.2 Generating Pseudo Random Numbers

A digital computer cannot really generate genuinely random numbers nor throw dices. Though, it can generate series of so-called pseudo random numbers in the (0, 1) interval by applying a deterministic algorithm to an initial positive integer number called the seed. If one knows this seed and the algorithm all drawings are perfectly predictable, but if not, they have the appearance of IID drawings from the uniform distribution over the 0-1 interval. If one just knows the seed but not the algorithm one cannot predict the series, but by using the algorithm with the same seed again one can reproduce the same pseudo random series whenever desired. This comes in handy when one wants to compare alternative methods under equal circumstances.

The algorithm for producing random numbers is usually of the following simple type. Let $z_0 > 0$ be the positive integer seed, then pseudo IID $U(0,1)$ drawings $(\eta_1, \eta_2, ...)$ follow from the iterative scheme

$$\left.\begin{aligned} z_i &= (\psi z_{i-1} + \alpha) \div m \\ \eta_i &= z_i/m, \end{aligned}\right\} \quad i = 1,2,\ldots \tag{1.1}$$

where integer m is called the modulus, ψ is the multiplier, and α the increment. The operation $\div$ (often denoted as *mod*) means here that z_i equals the remainder of dividing $\psi z_{i-1} + \alpha$ by m. In at most m steps the series of values z_i and thus η_i will repeat itself. Hence, m should be large, preferably as large as the largest integer on the computer, say $2^{31} - 1$. The choice of the value of ψ is crucial for the quality of the series too, but α is less relevant and is often set at zero. Testing the adequacy of random number generators is an art of its own. We will simply trust the default versions available in the computer package that we use.

The CDF of $\eta_i \sim U(0,1)$ is

$$F_U(\eta) \equiv \Pr(\eta_i \leq \eta) = \begin{cases} 0, & \eta < 0 \\ \eta, & 0 \leq \eta \leq 1 \\ 1, & \eta > 1. \end{cases} \tag{1.2}$$

By appropriately transforming IID drawings $\eta_i \sim U(0,1)$ one can obtain IID drawings ζ_i from any other type of distribution D with strictly increasing CDF given by $F_D(\zeta)$, with $\zeta \in \mathbb{R}$. Consider $\zeta_i = F_D^{-1}(\eta_i)$. This yields $F_D(\zeta_i) = \eta_i \sim U(0,1)$. So

$$\Pr(\zeta_i \leq \zeta) = \Pr(F_D(\zeta_i) \leq F_D(\zeta)) = \Pr(\eta_i \leq F_D(\zeta)) = F_D(\zeta) \tag{1.3}$$

indeed. Hence, generating $\zeta_i = F_D^{-1}(\eta_i)$ yields a series of IID pseudo random drawings of ζ_i, $i = 1, 2, \ldots$. This does not work out well when distribution D has a CDF that has no closed form for its inverse, as is the case for the (standard) Normal distribution. However, relatively simple alternative transformation techniques are available for that situation, see for instance Fishman (2006).

1.2.1 Drawings from $U(0,1)$

We shall now use EViews[1] to illustrate the above and interpret some realizations of series of generated random drawings. At the same time

[1] Check the EViews reference guide to find out about any particulars on the random number generator that your version of EViews uses. All results to follow were obtained by EViews 7. Earlier versions use a different random number generator and therefore do not yield results fully similar to those presented here.

we shall learn how to use the programming facilities of EViews. Consider the following EViews program:

```
'mcs11.prg:  Drawings from U(0,1)
!n=1000
workfile f:\MCS\mcs11.wf1 u 1 !n
rndseed 9876543210
genr eta=rnd
```

The first line of this program[2] (and all programs to follow) starts with the character "'" meaning that it just contains a comment and (without any computational consequences) exhibits the name (with extension prg) and purpose of the program. Names of integer or real variables have to be preceded by the character "!". By !n we identify sample size. In the third line we identify a new workfile and its location (map or folder); for clarity we give this workfile (which has extension wf1) the same name as the program. The parameter "u" indicates that the observations are "undated" (not associated with a particular year or quarter) and next it is indicated that their range will run form 1 to !n. In the fourth line we provide a specific but arbitrary integer seed value for the random number generator, and in the final line a variable eta is generated of n IID drawings from $U(0,1)$. After running this program in an EViews session one can manipulate and analyze the data series eta stored in the workfile as one wishes, either by using standard EViews commands or by running another program operating on this workfile.

Figure 1.1 presents the histograms, as produced by EViews, obtained from running program mcs11 first for $n = 1{,}000$ and next for $n = 1{,}000{,}000$. Both these histograms establish empirical probability density functions of the $U(0,1)$ distribution. Both deviate from the rectangular actual population PDF, and obviously and visibly the one with larger n is more accurate. The value of mean is calculated according to

$$\bar{\eta}_n \equiv \frac{1}{n}\sum_{i=1}^{n}\eta_i. \tag{1.4}$$

[2] All programs are available in a zipped file mcs.zip. These programmes suppose that you have access to a drive f:\ with folder MCS. Of course, the path f:\MCS\ can be changed in whatever is more convenient.

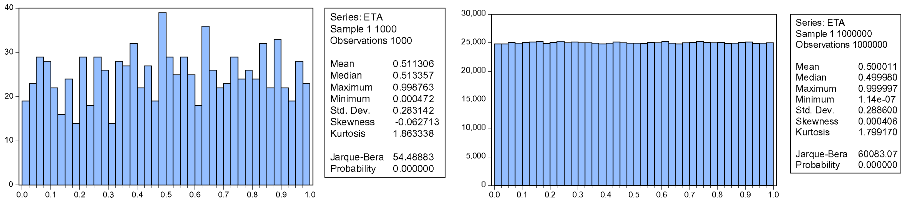


Fig. 1.1 Drawings from $U(0,1)$ for $n = 1{,}000$ and $n = 1{,}000{,}000$.

For both sample sizes these are pretty close to $E(\eta) = 0.5$. The reason is that, when the η_i establish a series of $U(0,1)$ drawings indeed, the sample average $\bar{\eta}_n$ is an unbiased estimator of $E(\eta)$, because

$$E(\bar{\eta}_n) = E\frac{1}{n}\sum_{i=1}^{n}\eta_i = \frac{1}{n}\sum_{i=1}^{n}E(\eta_i) = \frac{n}{n}E(\eta) = 0.5. \tag{1.5}$$

Of course, the actual deviation of mean from 0.5 is associated with its standard deviation. We find

$$\begin{aligned} \mathrm{Var}(\bar{\eta}_n) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\eta_i\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n}\eta_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(\eta_i) = \frac{1}{n}\mathrm{Var}(\eta) = \frac{1}{12n}, \end{aligned} \tag{1.6}$$

where the third equality follows from the independence of the drawings, yielding $\mathrm{Cov}(\eta_i,\eta_j) = 0$ for $i \neq j$, and the final one from using general results on the $U(a,b)$ distribution. Regarding its first four centered moments we have $\mu_1^c \equiv E[\eta - E(\eta)] = 0$, $\mu_2^c \equiv \sigma_\eta^2 \equiv E[\eta - E(\eta)]^2 = (b-a)^2/12$, $\mu_3^c \equiv E[\eta - E(\eta)]^3 = 0$, and $\mu_4^c \equiv E[\eta - E(\eta)]^4 = (b-a)^4/80$, respectively. Since $\sqrt{\mathrm{Var}(\bar{\eta}_n)} = (12n)^{-1/2} = 0.0091$ and 0.00029, for $n = 1{,}000$ and $n = 1{,}000{,}000$ respectively, we can understand why the Mean value is much closer to 0.5 in the larger sample. Note, however, that there will be values of the seed for which the sample mean for $n = 1{,}000$ is closer to 0.5, because it is not the deviation itself that will be larger at $n = 1{,}000$ than at $n = 1{,}000{,}000$, but for a smaller sample size the probability is larger that the deviation will be larger than a particular value. For similar reasons it is understandable that the estimate of the median is more accurate for the larger n.

The value of Std. Dev. (which is actually what we would usually call the standard error, because it is the estimated standard deviation) mentioned next to the histograms is obtained as the square root of

$$\hat{\sigma}_\eta^2 \equiv \frac{1}{n-1}\sum_{i=1}^{n}(\eta_i - \bar{\eta}_n)^2. \tag{1.7}$$

Both estimates are pretty close to the standard deviation of the $U(0,1)$ distribution, which is $\sqrt{1/12} = 0.2887$. This is again no surprise, because

$$E(\hat{\sigma}_\eta^2) = \sigma_\eta^2 = \mathrm{Var}(\eta) = 1/12. \tag{1.8}$$

Note that the definitions of skewness and kurtosis are $\mu_3^c/(\mu_2^c)^{3/2}$ and $\mu_4^c/(\mu_2^c)^2$ respectively, so the population values for the $U(0,1)$ distribution are 0 and $144/80 = 1.8000$, respectively. Again we find that the estimates obtained from the two samples are pretty close to their population values, and they are closer for the larger sample size. The improvements with n are due to the fact that the corresponding estimators do have a variance of order $O(n^{-1})$.

1.2.2 Drawings from $N(0,1)$

Next we adapt the program as follows:

```
'mcs12.prg  Drawings from N(0,1)
!n=1000
workfile f:\MCS\mcs12.wf1 u 1 !n
rndseed 9876543210
genr zeta=nrnd
```

This yields for $n = 1{,}000, 1{,}000{,}000$ the histograms of Figure 1.2. The results on these samples of $N(0,1)$ drawings can be analyzed in a similar way as we did for $U(0,1)$.[3] Here, however, we have $\mu_1 = 0$, $\mu_2 = 1, \mu_3 = 0$, and $\mu_4 = 3$.

[3] Note that $(0,1)$ indicates the domain for the uniform distribution, whereas it refers to expectation and variance in case of the normal.

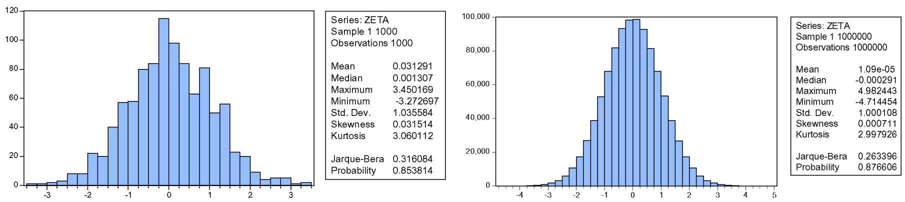


Fig. 1.2 Drawings from $N(0,1)$ for $n = 1{,}000$ and $n = 1{,}000{,}000$.

Note that the phenomenon that the values of mean converge to zero for increasing values of n illustrate the simplest form of the LLN (law of large numbers). Since both η_i and ζ_i are IID and have finite moments the sample average converges to the population mean (expectation). The same holds for the uncentered sample averages of powers of η_i and ζ_i which explains, upon invoking Slutsky's theorem, the convergence of their nonlinear transformations Std. Dev., Skewness and Kurtosis. See Appendix B for more details on the various tools (the notation $O(n^{-1})$, LLN, CLT, Slutsky) for asymptotic analysis.

1.3 LLN and Classic Simple Regression

Both in MCS and in econometrics the LLN plays a central role. Therefore, to understand its properties better, we will now provide some illustrations on the workings of the LLN in the context of a very simple regression model (and also on its limitations in Exercise 9). We consider the model with just one single exogenous regressor and start with the case where the observations are IID, and not necessarily normal. For $i = 1, \ldots, n$ the DGP (data generating process) is

$$y_i = \beta x_i + u_i, \quad x_i \sim IID(\mu_x, \sigma_x^2), \quad u_i \mid x_1, \ldots, x_n \sim IID(0, \sigma_u^2), \tag{1.9}$$

where μ_x, σ_x^2, and σ_u^2 are all finite. A well-known asymptotic (for $n \to \infty$) result in econometric theory is that in this model, due to the LLN, $\operatorname{plim} n^{-1} \sum_{i=1}^n x_i^2 = \lim n^{-1} \sum_{i=1}^n E(x_i^2) = \sigma_x^2 + \mu_x^2$ and $\operatorname{plim} n^{-1} \sum_{i=1}^n x_i u_i = \lim n^{-1} \sum_{i=1}^n E(x_i u_i) = 0$, since by the LIE (Law of Iterated Expectations) $E(x_i u_i) = E[E(x_i u_i \mid x_i)] = E[x_i E(u_i \mid x_i)] = E(0) = 0$. Employing Slutsky, we now find consistency for the OLS estimator, because

$$\begin{aligned} \operatorname{plim} \hat{\beta} &= \operatorname{plim} \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \\ &= \beta + \frac{\operatorname{plim} n^{-1} \sum_{i=1}^{n} x_i u_i}{\operatorname{plim} n^{-1} \sum_{i=1}^{n} x_i^2} = \beta + \frac{0}{\sigma_x^2 + \mu_x^2} = \beta. \end{aligned} \tag{1.10}$$

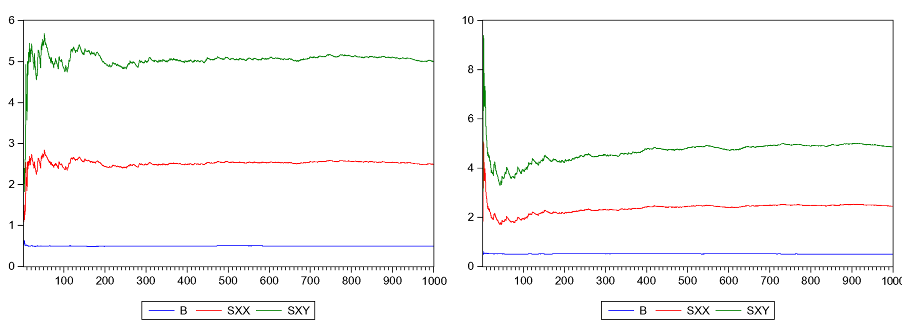
Of course, on a computer we cannot fully mimic the situation $n \to \infty$, but the analytic phenomena just discussed are nevertheless convincingly (especially when you increase the value of nmax) illustrated by the following program.

```
'mcs13.prg: LLN in simple IID regression
!nmax=1000
workfile f:\MCS\mcs13.wf1 u 1 !nmax
!beta=0.5
!mux=1
!sigx=2
!sigu=0.2
rndseed 9876543210
genr x=!mux + !sigx*(rnd - 0.5)/@sqrt(1/12)
genr u=!sigu*(rnd - 0.5)/@sqrt(1/12)
genr y=!beta*x + u
stom(x,vecx)
stom(y,vecy)
matrix (!nmax,3) results
!sumxy=0
!sumxx=0
for !n=1 to !nmax
  !sumxy=!sumxy+vecx(!n)*vecy(!n)
  !sumxx=!sumxx+vecx(!n)^2
  results(!n,1)=!sumxy/!n
  results(!n,2)=!sumxx/!n
  results(!n,3)=!sumxy/!sumxx
next
results.write f:\MCS\mcs13results.txt
read f:\MCS\mcs13results.txt sxx sxy b
```

In this program we chose $\beta = 0.5$, $x_i \sim IID(1, 2^2)$, $u_i \sim IID(0, 0.2^2)$ and both x_i and u_i are uniformly distributed and mutually independent. By the command "stom" we transform a generated series into a vector, which then enables to program expressions involving its individual elements. This allows to calculate both $n^{-1}\sum_{i=1}^n x_i^2$ and $n^{-1}\sum_{i=1}^n x_i y_i$ for $n = 1, \ldots, 1000$, and also their ratio. The results are stored in a matrix called results, which has nmax rows and 3 columns. In the two final lines of the program this 1,000 × 3 matrix is first written to a text file and from this the three columns are added as variables to the workfile mcs13.wf1 under the names sxy, sxx and b. These can then easily be analyzed further by EViews.

Figure 1.3 presents the graphs of sxy, sxx, and b as obtained for two different integer seed values. For small values of n the random nature of the three depicted statistics is apparent from the diagram, but they gradually converge for increasing n (see Exercise 6), and ultimately for $n \to \infty$, irrespective of the value of rndseed used, they assume their deterministic population values which are $\sigma_x^2 + \mu_x^2 = 5$, $\operatorname{plim} n^{-1}\sum_{i=1}^n x_i y_i = \beta \operatorname{plim} n^{-1}\sum_{i=1}^n x_i^2 = 2.5$, and $\beta = 0.5$ respectively. In fact, due to the correlation between sxy and sxx we note that the convergence of their ratio is much faster than that of sxy and sxx individually.

The LLN does not require that the observations in the regression are IID; they only have to be asymptotically uncorrelated. We will now verify the effects of first-order serial correlation in both the



rndseed = 9876543210 *rndseed* = −9876543210

Fig. 1.3 IID data; second moments (uncentered) and $\hat{\beta}$ for $n = 1, \ldots, 1{,}000$.

regressor x_i and the disturbances u_i. Because EViews generates variables recursively, in program msc14.prg variables $x_i^* \sim IID(0,\sigma_x^2)$ and $u_i^* \sim IID(0,\sigma_u^2)$ are actually generated first (but already stored in x_i and u_i) for $i = 1,\ldots,n\max$ and then the program calculates $u_1 = u_1^*$ and (for $i > 1$) $u_i = \rho_u u_{i-1} + (1-\rho_u^2)^{1/2}u_i^*$ and similarly for x_i, to which finally μ_x is added.

```
'mcs14.prg: LLN in simple non-IID regression
!nmax=1000
workfile f:\MCS\mcs14.wf1 u 1 !nmax
!beta=0.5
!mux=1
!sigx=2
!rhox=0.8
!rrhox=@sqrt(1-!rhox^2)
!sigu=0.2
!rhou=0.4
!rrhou=@sqrt(1-!rhou^2)
rndseed 9876543210
genr x=!sigx*(rnd - 0.5)/@sqrt(1/12)
genr u=!sigu*(rnd - 0.5)/@sqrt(1/12)
smpl 2 !nmax
genr x=!rhox*x(-1)+!rrhox*x
genr u=!rhou*u(-1)+!rrhou*u
smpl 1 !nmax
genr x=!mux + x
genr y=!beta*x + u
stom(x,vecx)
stom(y,vecy)
matrix (!nmax,3) results
!sumxy=0
!sumxx=0
for !n=1 to !nmax
  !sumxy=!sumxy+vecx(!n)*vecy(!n)
  !sumxx=!sumxx+vecx(!n)^2
  results(!n,1)=!sumxy/!n
```

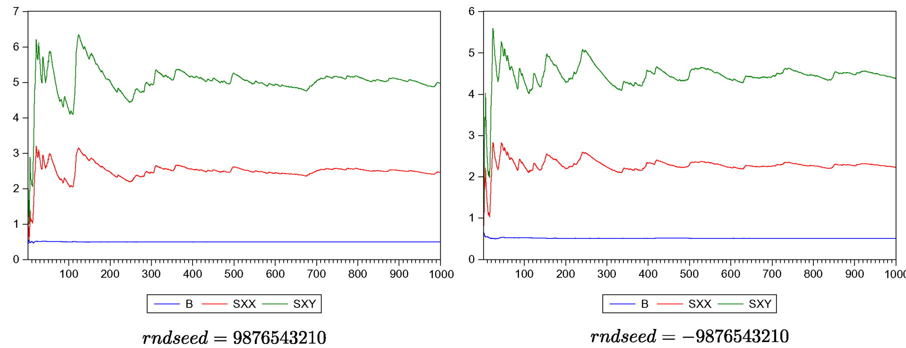


Fig. 1.4 Non-IID data; second moments (uncentered) and $\hat{\beta}$ for $n = 1, \ldots, 1{,}000$.

```
  results(!n,2)=!sumxx/!n
  results(!n,3)=!sumxy/!sumxx
next
results.write f:\MCS\mcs14results.txt
read f:\MCS\mcs14results.txt sxx sxy b
```

Figure 1.4 illustrates that OLS is also consistent in a model where either the observations of the regressors or those of the disturbances (or both) are serially correlated, but more and more uncorrelated at greater distance, provided that regressors and disturbances are contemporaneously uncorrelated. One can derive

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} x_i u_i\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n} x_i u_i\right)^2 = \frac{1}{n^2}E\left(\sum_{i=1}^{n} x_i u_i\right)^2 \\ &= \frac{\sigma_x^2\sigma_u^2}{n}\left(\frac{1+\rho_x\rho_u}{1-\rho_x\rho_u} - 2\frac{\rho_x\rho_u}{n}\frac{1-(\rho_x\rho_u)^n}{(1-\rho_x\rho_u)^2}\right),\end{aligned}$$

from which it follows that the convergence for the numerator, although of the same rate in n, is nevertheless slower for positive ρ_x and ρ_u than in the IID case, where $\rho_x = \rho_u = 0$. This is clearly illustrated by the diagrams.

In both programs mcs13.prg and mcs14.prg the regressor variable is strongly exogenous, because $E(u_i \mid x_1, \ldots, x_n) = 0\ \forall i$, and the disturbances are homoskedastic. Note that in establishing the consistency of OLS we only used $E(u_i \mid x_i) = 0$. So, neither serial correlation nor heteroskedasticity of the disturbances would spoil this result. Also weak exogeneity or predeterminedness of the regressors, where $E(u_i \mid x_1,\ldots,x_i) = 0$, still yields consistency of OLS.

1.4 CLT and Simple Sample Averages

The conclusions that we will draw on the basis of MCS studies will rely heavily, as far as their accuracy is concerned, on a very straight-forward application of the simplest version of the Central Limit Theorem. In MCS we will often approximate the sample average of IID observations generated from a usually unknown distribution by the normal distribution. In fact, we will standardize the sample average and approximate the outcome with the standard normal distribution. This approximation is perfect when it concerns a sample of NIID observations or when the sample is infinitely large, but it will involve approximation errors when the sample observations have a nonnormal distribution and the sample size is finite. In the illustration to follow we will examine the quality of the approximation for a few different nonnormal distributions and for various finite sample sizes.

Program mcs15.prg calculates sample averages from a sample of size n for five different IID variables. These variables are $(i = 1,\ldots,n)$: $z_i \sim N(0,1)$, $v_i \sim N(\mu_v,\sigma_v^2)$, $u_i \sim U(a,b)$, $x_i \sim \chi^2(2)$ and w_i, where the latter is a mixture of independent $\chi^2(1)$ and $\chi^2(2)$ variables. The generation of samples of size n of these variables, and the calculation of the sample averages is replicated R times. Running the program results in two workfiles. Every replication workfile mcs15.wf1 contains the n observations on the five variables, and after termination of the program these are their realizations in the final replication. Workfile mcs15res.wf1 contains variables of R observations. These are the generated sample averages and also their rescaled versions, in deviation from their expectation and divided by the standard deviation of the sample average. Of course, the latter equals the standard deviation of the individual elements divided by $\sqrt{n}$ (prove this yourself!). The CLT implies that for $n \to \infty$ the rescaled expressions should be indistinguishable from drawings from the $N(0,1)$ distribution. By the program we will examine how close we get to that when n is 10, 100, or 1,000.

In fact, we run the program first for $n = 1$, not because we expect the CLT to have much to say then, but because this is a straight-forward way to obtain the histograms of R drawings from the distributions of the five different random variables from this program.

```
'mcs15.prg:  Sample averages and the CLT
!n=10
workfile f:\MCS\mcs15.wf1 u 1 !n
!muv=1
!sigv=3
!a=5
!b=15
rndseed 9876543210
!R=10000
matrix (!R,5) simres
for !rep=1 to !R
   genr z=nrnd
   genr v=!muv + !sigv*z
   genr u=!a + (!b-!a)*rnd
   genr x=@rchisq(2)
   genr p=rnd>0.75
   genr w=p*(x+3) - (1-p)*(z^2+2)
   simres(!rep,1)=@mean(z)
   simres(!rep,2)=@mean(v)
   simres(!rep,3)=@mean(u)
   simres(!rep,4)=@mean(x)
   simres(!rep,5)=@mean(w)
next
simres.write f:\MCS\mcs15res.txt
workfile f:\MCS\mcs15res.wf1 u 1 !R
read f:\MCS\mcs15res.txt meanz meanv meanu meanx meanw
genr rescaledmeanz=@sqrt(!n)* meanz
genr rescaledmeanv=@sqrt(!n)*(meanv - !muv)/!sigv
genr rescaledmeanu=@sqrt(!n)*(meanu - (!a + !b)/2)/((!b - !a)/@sqrt(12))
genr rescaledmeanx=@sqrt(!n)*(meanx - 2)/2
genr rescaledmeanw=@sqrt(!n)*(meanw + 1)/@sqrt(14.5)
```

In Figure 1.5 histograms are shown of 10,000 drawings from $v \sim N(1, 3^2)$, $u \sim U(5, 15)$, $x \sim \chi^2(2)$ and w, where the latter two are non-symmetric and the last one is clearly bimodal. Note that u has thin tails (low kurtosis), that x is skew to the right and has very high kurtosis, whereas w also has positive skewness and kurtosis larger than for the normal distribution. The Jarque–Bera test notes indeed that the latter three distributions are seriously nonnormal.

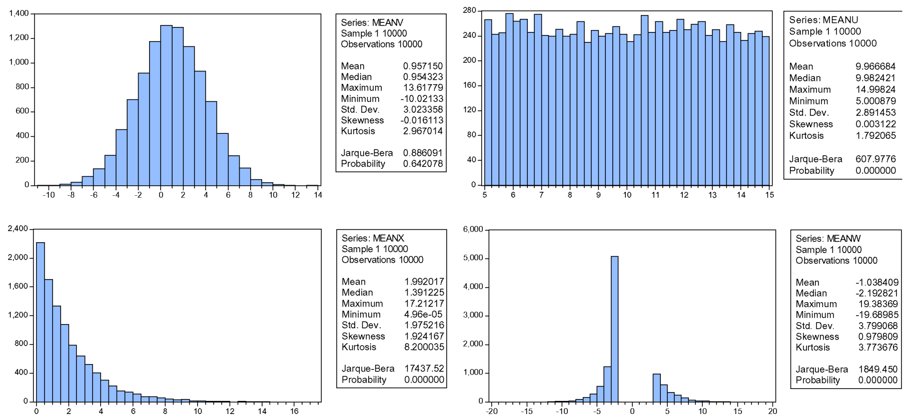


Fig. 1.5 Histograms of 10,000 drawings of v, u, x, and w.

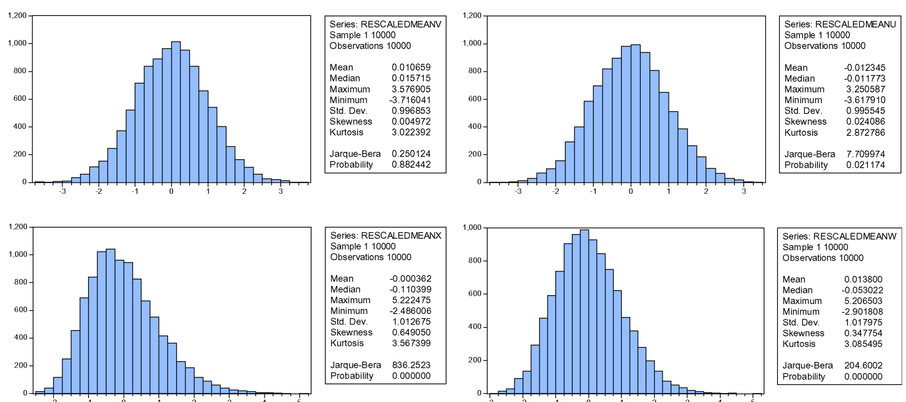


Fig. 1.6 Distribution of rescaled sample averages for $n = 10$.

In Figure 1.6, $n = 10$ and histograms of the rescaled sample averages are presented. From the histograms of 10,000 drawings it is obvious that even at $n = 10$ these sample averages already started to converge toward the normal. This is self-evident for rescaledmeanv, because these are drawings from the standard normal distribution for any n. The symmetry of the uniform distribution implies that rescaledmeanu has skewness very close to zero for any n, and at $n = 10$ the kurtosis is already such that the normality hypothesis, although actually invalid, is not rejected at the 1% level. Also for the sample averages of the x

and w distributions the skewness and the kurtosis are already much closer to zero and three respectively, although still such that normality is strongly rejected by the Jarque–Bera test.

Figure 1.7 contains results for $n = 100$ and shows that averaging has completely removed the underlying nature of the uniform drawings and of the bimodality of the w_i drawings, but the skew nature of the x_i and w_i distributions is still emerging from the averages of samples of this size.

Continuing this and taking $n = 1{,}000$ it is shown in Figure 1.8 that the sample averages are again closer to normal. These illustrations

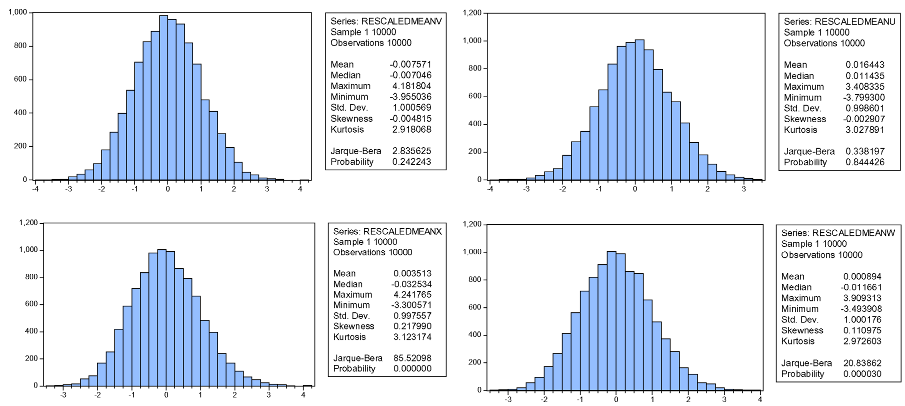


Fig. 1.7 Distribution of rescaled sample averages for $n = 100$.

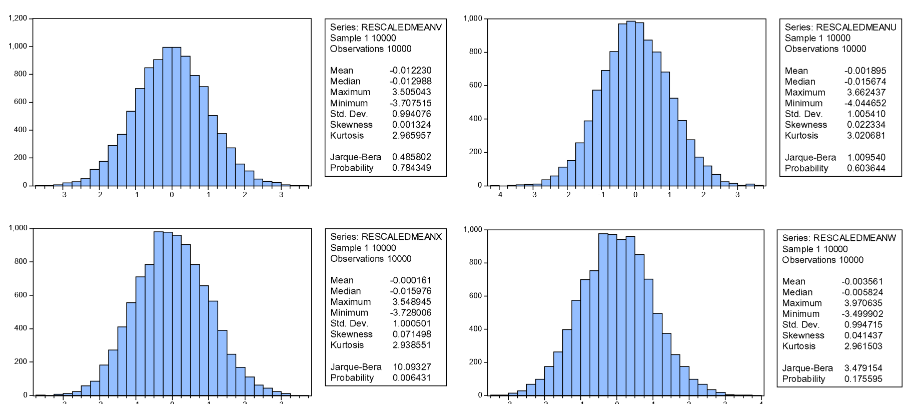


Fig. 1.8 Distribution of rescaled sample averages for $n = 1{,}000$.

clearly demonstrate that both the degree of nonnormality of the underlying random variables and the size of the sample over which the average is taken jointly determine the accuracy of the normal approximation. Seriously nonnormal distributions are shown to have sample averages that are distributed closely to (although possibly still significantly different from) normal when the sample is as large as 1,000. In fact, when it comes to the accuracy of the MCS inferences to be developed in the following sections, it will be argued that it does not matter that much whether or not the distribution of a sample average as such is very accurately approximated by the normal distribution, but only whether its tail probabilities and thus its quantiles in the tail areas conform closely to those of the normal. We will examine that in more detail later, and also other aspects that determine the accuracy of MCS inference.

In most of the simulated above results, we obtained information on relatively simple statistics for which many of their typical properties, especially their moments, can be derived analytically. So, the simulation results merely serve as a specific numerical illustration of already fully understood more general characteristics. This will also be the case in the more involved MCS illustrations on particular parametrizations of the standard normal linear regression model in the next section. The distribution of their relevant statistics can be derived analytically, so there is no genuine need for simulating them other than illustration. However, such results help to appreciate that MCS results lack generality, are nonanalytical but numerical, and are random as well. Therefore, they are both very specific and involve inaccuracies, and only after fully understanding the nature and magnitude of these inaccuracies we will move on and apply MCS in a range of situations where the true underlying properties of estimators and tests are mostly unknown and the estimated numerical results from MCS establish, next to their analytic asymptotic approximations, our only guidebook.

Exercises

1. Consider (1.8). Prove $E(\hat{\sigma}_\eta^2) = \mathrm{Var}(\eta)$. Also, from $F_U(\eta)$ in (1.2), find the density of $U(0,1)$ and derive $\mathrm{Var}(\eta) = 1/12$.

2. Explain the high values of the Jarque–Bera statistics (consult the EViews Help facility for more details on this test) in Figure 1.1 and argue why these lead to low p-values. General warning: Never mistake the p-value of a test statistic as expressing the probability that the null hypothesis is true, because it simply expresses the probability (according to its asymptotic null distribution) that the test statistic may assume a value as extreme as it did (hence, supposing that the null hypothesis is true).
3. Consider the estimator $\hat{\sigma}^2_\eta$ of (1.7) where $\eta_i \sim UIID(0,1)$. Derive $\text{Var}(\hat{\sigma}^2_\eta)$ and show that it is $O(n^{-1})$.
4. Explain the moderate values of the Jarque–Bera statistics in Figure 1.2 and their corresponding high p-values.
5. Run similar programs as mcs11.prg and mcs12.prg with different values of seed and similar (and also larger) values of n and explain your findings.
6. Run programs mcs13.prg and mcs14.prg for $n = 100,000$ upon changing rndseed and possibly also choosing different values for the parameters μ_x, σ_x, and σ_u. Compare the sample equivalents of μ_x, σ_x, and σ_u with their population values. Also interpret the correlogram of the u and x series.
7. Run program mcs15.prg for $n = 1{,}000$ to replicate the results of Figure 1.7. Consider the histogram of variable meanw. Derive analytically that $E(w) = -1$ and $\text{Var}(w) = 14.5$. Explain why the histograms of meanw and rescaledmeanw show similar values for skewness and kurtosis. Give command genr rejw=abs(rescaledmeanw)>1.96 in the command window of EViews and examine the histogram of variable rejw. Note that, although the distribution of rescaledmeanw does not correspond to the standard normal in all aspects, the probability that a drawing is larger in absolute value than the 2.5% quantile of the standard normal does not seem to differ much from 5%.
8. In Figure 1.8 normality of rescaledmeanu is not rejected. What will be the effect on this finding from running the program for $n = 1{,}000$ with a much larger value of R? And

the same question for any R but a much larger value of n? Try and explain.

9. Run program mcs16.prg (given below) and choose various different rndseed values while keeping the number of degrees of freedom v fixed and examine the histogram of the generated series of random variables for $v = 3, 2, 1$. Note that for a random variable following the *Student*(v) distribution only the moments up to $v - 1$ exist. Its higher-order moments are defined by an integral which is infinite, because the tails of its density function are relatively fat. So, for $v = 3$ the skewness and kurtosis estimates do not converge (the LLN does not apply!). Irrespective of the value of n they yield different outcomes when rndseed is changed. For $v = 2$ the same happens with the standard deviation estimate, and for $v = 1$ (this is the Cauchy distribution) also the sample average has no deterministic population equivalent. Note, however, that the median does converge to zero for n large, irrespective of the values of v and of rndseed.

```
'mcs16.prg       Drawings from Student(v)
!n=10000
workfile f:\MCS\mcs16.wf1 u 1 !n
rndseed 9876543210
!v=3
genr studentv=@rtdist(!v)
```

References

Arellano, M. and S. R. Bond (1991), 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations'. *The Review of Economic Studies* **58**, 277–297.

Brown, B. W. and W. K. Newey (2002), 'Generalized method of moments, efficient bootstrapping, and improved inference'. *Journal of Business and Economic Statistics* **20**, 507–517.

Bun, M. J. G. and J. F. Kiviet (2006), 'The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models'. *Journal of Econometrics* **132**, 409–444.

Cramér, H. (1946), *Mathematical Methods of Statistics*. Princeton.

David, H. A. (1981), *Order Statistics*. Wiley.

Davidson, R. and J. G. MacKinnon (1993), *Estimation and Inference in Econometrics*. New York: Oxford University Press.

Davidson, R. and J. G. MacKinnon (1998), 'Graphical methods for investigating the size and power of hypothesis tests'. *The Manchester School* **66**, 1–26.

Davidson, R. and J. G. MacKinnon (2004), *Econometric Theory and Methods*. Oxford University Press.

Davidson, R. and J. G. MacKinnon (2006), 'Bootstrap methods in econometrics'. In: T. C. Mills and K. Patterson (eds.): *Palgrave*

Handbook of Econometrics, vol. 1*, Econometric Theory*. Basingstoke, Palgrave Macmillan, pp. 812–838.

Doornik, J. A. (2006), 'The role of simulation in econometrics'. In: T. C. Mills and K. Patterson (eds.): *Palgrave Handbooks of Econometrics*, vol. 1*, Econometric Theory*. Basingstoke, Palgrave Macmillan, pp. 787–811.

Dufour, J.-M. and L. Khalaf (2001), 'Monte Carlo test methods in econometrics'. In: B. Baltagi (ed.): *Companion to Theoretical Econometrics*. Oxford, UK: Blackwell, pp. 494–519.

Edgerton, D. L. (1996), 'Should stochastic or non-stochastic exogenous variables be used in Monte Carlo experiments?'. *Economics Letters* **53**, 153–159.

Fishman, G. S. (2006), *A First Course in Monte Carlo*. Thomson.

Gourieroux, C. and A. Monfort (1996), *Simulation-Based Econometric Methods*. Oxford University Press.

Greene, W. H. (2012), *Econometric Analysis*. 7th international edition. Pearson.

Hendry, D. F. (1984), 'Monte Carlo experimentation in econometrics'. In: Z. Griliches and M. D. Intriligator (eds.): *Handbook of Econometrics*, vol. II. Amsterdam: Elsevier.

Hendry, D. F. (1995), *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F. and B. Nielsen (2007), *Econometric Modelling*. Oxford: Princeton University Press.

Horowitz, J. L. (2003), 'The bootstrap in econometrics'. *Statistical Science* **18**, 211–218.

Johnson, R. W. (2001), 'An introduction to the bootstrap'. *Teaching Statistics* **23**, 49–54.

Kiviet, J. F. (1985), 'Model selection test procedures in a single linear equation of a dynamic simultaneous system and their defects in small samples'. *Journal of Econometrics* **28**, 327–362.

Kiviet, J. F. (2007), 'Judging contending estimators by simulation: Tournaments in dynamic panel data models'. In: G. D. A. Phillips and E. Tzavalis (eds.): *The Refinement of Econometric Estimation and Test Procedures; Finite Sample and Asymptotic Analysis*. Cambridge University Press, pp. 282–318.

Kiviet, J. F. and J. Niemczyk (2011), 'Comparing the asymptotic and empirical (un)conditional distributions of OLS and IV in a linear static simultaneous equation'. To appear in*: Journal of Computational Statistics & Data Analysis*.

Kiviet, J. F. and G. D. A. Phillips (2012), 'Higher-order asymptotic expansions of the least-squares estimation bias in first-order dynamic regression models'. To appear in: *Journal of Computational Statistics & Data Analysis*.

MacKinnon, J. G. (2002), 'Bootstrap inference in econometrics'. *Canadian Journal of Economics* **35**, 615–645.

MacKinnon, J. G. (2006), 'Bootstrap methods in econometrics'. *Economic Record* **82**, s2–s18.

Metropolis, N. and S. Ulam (1949), 'The Monte Carlo method'. *Journal of the American Statistical Society* **44**, 335–341.

Nelson, R. N. and R. Startz (1990), 'Some further results on the exact small sample properties of the instrumental variable estimator'. *Econometrica* **58**, 967–976.

Stigler, S. M. (1999), *Statistics on the Table; The History of Statistical Concepts and Methods*. Cambridge MA: Harvard University Press.

Student (1908), 'On the probable error of the mean'. *Biometrika* **6**, 1–25.

Wagner, H. M. (1958), 'A Monte Carlo study of estimates of simultaneous linear structural equations'. *Econometrica* **26**, 117–133.