GMM, Indirect Inference and Bootstrap

Exam

Winter 2015/2016

- Answer **all** of the following three exercises in either German or English.
- Hand in your solutions before March, 8 2016 at 12:00.
- Please e-mail the solutions to mutschler@statistik.tu-dortmund.de
- The solution files should contain your executable (and commented) R code and an additional documentation preferably as `pdf`, not `doc` or `docx`.
- All students must work on their own, please also give your student ID number.
- It is advised to regularly check the homepage in case of urgent updates.
- If there are any questions, do not hesitate to contact Willi Mutschler.

1 Test for the Zipf index of city size distributions

It is well known that the population size distribution of large cities can be approximated by the Zipf distribution which is a special case of the Pareto distribution with tail index $\alpha = 1$. Suppose, $X_1 \geq X_2 \geq \ldots \geq X_n$ is a descendingly ordered sample of city sizes. In regional economics, the tail index α is often estimated from the regression

$$\ln(i) = c - \alpha \ln X_i + u_i$$

where i is the rank of the city and X_i its size, the intercept parameter c is of no interest.

(a) Comment briefly on the properties of OLS for this regression. Do you expect the ordinary t-test to work properly? [Maximum of 4 sentences]

Ordered samples from the Zipf distribution (i.e. from the Pareto distribution with true tail index $\alpha = 1$) are generated by the command `x <- sort(exp(rexp(n)),decreasing=TRUE)` where `n` is the sample size.

(b) Simulate $R = 1000$ samples of size $n = 20$. Perform an OLS regression and plot the distribution of $\hat{\alpha}$.

(c) An important hypothesis is $H_0 : \alpha = 1$ against $H_1 : \alpha \neq 1$. Simulate $R = 1000$ samples of size $n = 20$ and plot the distribution of the test statistic

$$T = \frac{\hat{\alpha} - 1}{SE(\hat{\alpha})}$$

Show that it is not t_{n-2}-distributed even though H_0 is true. Also, provide economic intuition behind this hypothesis.

(d) Find the critical values of a **parametric** LM-type bootstrap test of $H_0 : \alpha = 1$ with 1000 bootstrap repetitions. Compare this bootstrap approach to part (c).

2 Finite sample properties for the Gamma model

Let y_t have an iid gamma distribution for $t = 1, \dots T$ with shape parameter $\alpha_0 = \{1, 3, 5\}$ and scale parameter $\beta_0 = 1$. Assume that β_0 is known.

(a) Investigate the finite sample bias and standard error of the following estimators of α_0, for sample sizes of $T = \{50, 100, 200\}$ with 1000 replications.

   (i) The maximum likelihood estimator.

   (ii) The GMM estimator based on the first moment of the gamma distribution.

   (iii) The GMM estimator based on the first two moments of the gamma distribution.

   Hints: `y <- rgamma(T,shape=a,scale=1)` *generates T draws from the gamma distribution with shape parameter a and scale parameter 1.* `lgamma(a)` *computes the logarithm of the gamma function at a. Since we are only interested in the finite sample properties of one parameter, use an optimizer for one-dimensional problems, see the help of* `optim`, `optimize`, `nlminb` *or any other optimization package for details. Use* `mean(y)` *as the starting value.*
   To get a nice overview, try to output your results for each sample size in a table:

Table 1: Bias and standard error given sample size T={50,100,200}

	$\hat{\alpha}_{ML}$		$\hat{\alpha}_{GMM1}$		$\hat{\alpha}_{GMM2}$	
α_0	Bias	S.D	Bias	S.D.	Bias	S.D
1	x.xxxx	x.xxxx	x.xxxx	x.xxxx	x.xxxx	x.xxxx
3	x.xxxx	x.xxxx	x.xxxx	x.xxxx	x.xxxx	x.xxxx
5	x.xxxx	x.xxxx	x.xxxx	x.xxxx	x.xxxx	x.xxxx

   Interpret your results.

(b) Compute the finite sample size and power of the t statistic $t = \frac{\hat{\alpha}-1}{sd(\hat{\alpha})}$ based on the three estimators in part (a) for sample size $T = 200$ and 1000 replications. Use parameter values of $\alpha_0 = \{1.00, 1.05, 1.10, 1.15, 1.20, 1.25, 1.30\}$.
   Hints: The size of the test is given for $\alpha_0 = 1$ and the power of the test is given for values $\alpha_0 > 1$. Count the number of times the absolute value of the computed t statistic is greater than its critical value of 1.96 (for nominal size 5%) and divide this by the number of repetitions. Output this similar to Table 2.

Table 2: Size and power for sample size T=200

α_0	$\lvert t_{ML}\rvert > 1.96$	$\lvert t_{GMM1}\rvert > 1.96$	$\lvert t_{GMM2}\rvert > 1.96$
1.00	x.xxxx	x.xxxx	x.xxxx
1.05	x.xxxx	x.xxxx	x.xxxx
1.10	x.xxxx	x.xxxx	x.xxxx
1.15	x.xxxx	x.xxxx	x.xxxx
1.20	x.xxxx	x.xxxx	x.xxxx
1.25	x.xxxx	x.xxxx	x.xxxx
1.30	x.xxxx	x.xxxx	x.xxxx

   Interpret your results.

(c) Suppose that the data-generating process is now an exponential distribution with scale parameter equal to α_0, whereas the estimators in part (a) are still based on the gamma distribution. Redo part (a) for sample sizes of $T = \{200, 400\}$ and discuss the effects of misspecification on the sampling properties of the three estimators.
   Hint: `y <- rexp(T,rate=1/a)` *generates T draws from the exponential distribution with scale parameter a.*

3 Stochastic volatility

An important area of the use of simulation estimation techniques in economics and finance is the estimation of continuous-time models, since, in general, maximum likelihood estimation is intractable for continuous-time processes. This is due to the fact, that to estimate such models it is necessary to have continuous data. However, in practice one observes data only at discrete points in time. The usual approach to resolve the difference in frequency between the model and data, is to discretize the model, thereby matching the frequency of the model to the data. The problem with this strategy is that discretizing a continuous-time model results in a misspecification of the functional form of the model, so that the parameters of the discrete model are, in general, not the parameters of the continuous-time model, resulting in biased estimators.

(a) Explain briefly how simulation estimation procedures such as *indirect inference* circumvent this misspecification problem (maximum of 4 sentences).

Consider the Black-Scholes model for the stock market. Assume that the price of a stock follows a geometric Brownian motion

$$dy_t = \mu y_t dt + \sigma_t y_t dW_t^y \tag{1}$$

where $t > 0$ denotes time and dt is the infinitesimal change in time. Furthermore, dy_t is the incremental change in the stock price y_t and μ is the expected return of the stock. W_t^y is a Wiener process such that $dW_t^y \overset{iid}{\sim} N(0, dt)$. Note that this model implies stochastic volatility, since $\sigma_t > 0$ is dependent on t. Assume that $ln(\sigma_t^2)$ follows a mean-reverting Ornstein-Uhlenbeck process

$$dln(\sigma_t^2) = \alpha(\kappa - ln(\sigma_t^2))dt + \gamma dW_t^\sigma \tag{2}$$

with $dln(\sigma_t^2)$ denoting the change in log volatility, κ its expected value, $\alpha > 0$ a sensitivity parameter and $\gamma > 0$ a diffusion parameter. W_t^σ is a Wiener process, such that $dW_t^\sigma \overset{iid}{\sim} N(0, dt)$. Furthermore, dW_t^y and dW_t^σ are assumed to be independent.

We would like to estimate the parameters $\theta = (\mu, \alpha, \kappa, \gamma)$ of this continuous-time model with stochastic volatility. Note that $ln(y_t)$ is observed discretely, whereas the volatility series $ln(\sigma_t^2)$ is latent, i.e. unobservable.

Equations (1) and (2) can be exactly discretized ($dt \approx \Delta t$ is a small number) with

$$ln(y_t) = a_y + b_y ln(y_{t-1}) + c_y u_t^y \tag{3}$$

$$ln(\sigma_t^2) = a_\sigma + b_\sigma ln(\sigma_{t-1}^2) + c_\sigma u_t^\sigma \tag{4}$$

where

$$a_y = \left(\mu - \frac{1}{2}\sigma_t^2\right)\Delta t \qquad b_y = 1 \qquad c_y = \sqrt{\sigma_t^2 \Delta t}$$

$$a_\sigma = \kappa\left(1 - b_\sigma\right) \qquad b_\sigma = e^{-\alpha \Delta t} \qquad c_\sigma = \gamma\sqrt{\Delta t}\sqrt{\frac{1 - b_\sigma^2}{2\alpha}}$$

and u_t^y and u_t^σ are independent standard normally distributed random variables.

(b) Write an R function that generates T discrete values for $ln(y_t)$. The function should:

(i) Simulate $n = \frac{T}{\Delta t}$ values for $ln(\sigma_t^2)$ and $ln(y_t)$ using equations (3) and (4).
Hint: To speed things up you can use the `filter` function.

(ii) Select every $\frac{1}{\Delta t}$-th value of $ln(y_t)$ and output this as a vector.

(iii) Set $\Delta t = 0.1$ and $T = 100$. Generate a sample with $\mu = 1$, $\alpha = 0.9$, $\kappa = 0$ and $\gamma = 1$. Save this as `lnytrue`. Comment on your choice of starting values.

(c) Continuous-time models are sometimes estimated by discretizing them in a crude way. The discretized version of (1) and (2) is

$$ln(y_t) - ln(y_{t-1}) = \mu + \sigma_t u_t^y \tag{5}$$

$$ln(\sigma_t^2) - ln(\sigma_{t-1}^2) = \alpha(\kappa - ln(\sigma_{t-1}^2)) + \gamma u_t^\sigma \tag{6}$$

with u_t^y and u_t^σ are independent standard normally distributed random variables. We need to derive indirect estimators for the parameters $\theta = (\mu, \alpha, \kappa, \gamma)$ given the auxiliary model in equations (5) and (6) that can be computed fast. Note, however, that we only have data for y_t, but not for σ_t. Therefore, do the following steps:

(i) Derive an unbiased estimator $\hat{\mu}$ for μ in (5) given data y_t.

(ii) Define $r_t = ln(y_t) - ln(y_{t-1}) - \mu$ and show that equations (5) and (6) can be manipulated to derive an expression

$$ln(r_t^2) = \delta_0 + \delta_1 ln(r_{t-1}^2) + \varepsilon_t \tag{7}$$

which is independent of σ_t. Furthermore, note that δ_0 and δ_1 are functions of α and κ, whereas ε_t is a function of γ and the error terms in (5) and (6).

(iii) Derive estimators for δ_0, δ_1 and the standard deviation of ε_t, $sd(\varepsilon_t)$, in equation (7), given data y_t and the unbiased estimator $\hat{\mu}$ for μ in (i), that can be computed fast. Note that this implicitly defines estimators for α, κ and γ.

(d) Load the dataset you generated in (b). Estimate the parameters $\theta = (\mu, \alpha, \kappa, \gamma)$ by indirect inference with the auxiliary model (5), (6) and (7). Your moment conditions are based on $\hat{\beta} = (\hat{\mu}, \hat{\delta}_0, \hat{\delta}_1, \widehat{sd(\varepsilon_t)})$ for the true dataset and the average of $\hat{\beta}_h = (\hat{\mu}_h, \hat{\delta}_{0,h}, \hat{\delta}_{1,h}, \widehat{sd(\varepsilon_t)_h})$ for $h = 1, \ldots, H$ simulated datasets. The indirect inference estimator is then defined as

$$\hat{\theta} = \underset{\theta}{argmin} \left\{ \left(\hat{\beta} - \overline{\hat{\beta}_h}\right)' W \left(\hat{\beta} - \overline{\hat{\beta}_h}\right) \right\}$$

with $\overline{\hat{\beta}_h} = \frac{1}{H} \sum_{h=1}^{H} \hat{\beta}_h$. Set $H = 10$ and use the identity matrix as the weight matrix W.

(e) Generate $R = 100$ new samples each with $\mu = 1$, $\alpha = 0.9$, $\kappa = 0$ and $\gamma = 1$. Repeat step (d) for each sample and save the estimators in a matrix. Compute the mean, standard deviation and root-mean-squared-error of your estimates.

(f) Comment on how your results would change if you used the optimal weight matrix instead of the identity matrix (maximum of 4 sentences).