Advanced Econometrics

Catalogue of Exercises - Status: 4.4.2024[1]

This catalogue contains 18 pages and consists of 45 tasks with a total of 332 points.

Exercise list

I would like to thank Roland Weigand and Stefan Rameseder for their input, important corrections and substantial suggestions for improvement.

1. Task (36 points) NLS: (**The Michaelis-Menten model**)

The Michaelis-Menten model is used, for example, to investigate clinical dose-response relationships. It is given by the following non-linear regression model:

$$y_t = \beta_1 + \frac{\beta_2 x_t}{x_t + \beta_3} + u_t, \quad u_t|x_t \sim IID(0, \sigma^2), \quad t = 1, 2, \ldots, n. \tag{1}$$

(a) (3 points) Provide $x_t(\boldsymbol{\beta})$ and $\boldsymbol{X}_t(\boldsymbol{\beta})$. Then calculate the second-order Taylor approximation of $x_t(\boldsymbol{\beta})$ around $\boldsymbol{\beta}_0$. This Taylor approximation is given by

$$f(\boldsymbol{x} + \boldsymbol{a}) \approx f(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})^T \boldsymbol{a} + \frac{1}{2}\boldsymbol{a}^T \boldsymbol{H}(\boldsymbol{x})\boldsymbol{a},$$

where $\boldsymbol{g}(\boldsymbol{w})$ denotes the gradient vector $\frac{\partial f}{\partial \boldsymbol{x}}\Big| \boldsymbol{x} = \boldsymbol{w}$ and $\boldsymbol{H}(\boldsymbol{w})$ denotes the Hessian matrix $\frac{\partial^2 f}{\partial \boldsymbol{x} \partial \boldsymbol{x}^T}\Big| \boldsymbol{x} = \boldsymbol{w}$ at point $\boldsymbol{w}$.

(b) (3 points) Compute the first two derivatives of the sum of squared errors $Q(\boldsymbol{\beta}) = \frac{1}{n}\sum_{t=1}^{n}(y_t - x_t(\boldsymbol{\beta}))^2$ with respect to $\boldsymbol{\beta}$. Represent these derivatives both in summation notation and in matrix notation.

You want to estimate $\boldsymbol{\beta}$ using a nonlinear method of moments estimator $\tilde{\boldsymbol{\beta}}$. This estimator is given by the moment condition $\mathbf{W}^T(\mathbf{y} - \mathbf{x}(\tilde{\boldsymbol{\beta}})) = \mathbf{0}$.

(c) (1 point) What dimension does $\mathbf{W}$ have?

(d) (2 points) A colleague suggests choosing $\mathbf{W}_t = (1;\ x_t;\ 5x_t)$. What do you reply?

(e) (2 points) For which values of $\boldsymbol{\beta}_0$ is the asymptotic identification violated for arbitrary instruments $\mathbf{W}_t$?

(f) (3 points) Outline briefly the derivation of the asymptotic distribution of $\tilde{\boldsymbol{\beta}}$ for a given instrument matrix $\mathbf{W}$ under appropriate assumptions.

(g) (2 points) Describe how you can construct an efficient estimator for $\boldsymbol{\beta}$ from (1) based on $\tilde{\boldsymbol{\beta}}$.

You want to perform nonlinear least squares estimation for the model (1). Let $\boldsymbol{x} = (x_1; x_2; \ldots; x_n)^T$. The objective function here is given by:

$$Q(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{x}) = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta}))^T(\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta})). \tag{3}$$

(h) (3 points) Show that under the usual assumptions and the additional assumption of a random sample, it holds that:

$$E\left[\frac{\partial^2 Q(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{x})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right] = 2E\left[\mathbf{X}_t(\beta_0)^T \mathbf{X}_t(\beta_0)\right] \quad \text{and}$$

$$\text{plim}\left(\frac{\partial^2 Q(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{x})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right) = 2\boldsymbol{E}\left[\mathbf{X}_t^T(\beta_0)\mathbf{X}_t(\beta_0)\right].$$

What changes if you consider $x_t, ; t = 1, \ldots, n$ as fixed (non-stochastic) values?

(i) (4 points) Let $\boldsymbol{\theta} := (\beta_1;;\beta_2)^T$. For a given (fixed) β_3, provide the estimator $\breve{\boldsymbol{\theta}}(\beta_3)$ that minimizes the sum of squared residuals (3). Then, provide the centered objective function

$$Q^z(\beta_3|\boldsymbol{y},\boldsymbol{x}) = Q\left((\breve{\boldsymbol{\theta}}(\beta_3),\ \beta_3)^T|\boldsymbol{y},\boldsymbol{x}\right).$$

(j) (4 points) Now, use the dataset `dose_data.csv`, where besides the administered dose `dose` of a medication, there is also a measure of its effect `response`. Explain the following R code, focusing on the practical utility of $Q^z(\beta_3;\boldsymbol{y},\boldsymbol{x})$.

```
data <- read.csv("dose_data.csv")

Qz <- function(b3,data){
est_lin <- lm(response~I(dose/(dose+b3)),data=data)
u <- est_lin$resid
sum(u^2)
}

b3_seq <- seq(-4,4,0.05)
Qz <- Vectorize(Qz,"b3")
plot(b3_seq, Qz(b3_seq, data=data),
  type="l",
  xlab=expression(beta[3]),
  ylab=expression(Q^z(beta[3])))

min_Qz <- optimize(Qz, interval=c(-3,3), data=data)
min_Qz
```

(k) (3 points) Calculate the nonlinear least squares estimator $\hat{\boldsymbol{\beta}}$ using the `nls()` function. Try different initial values for optimization, such as $\boldsymbol{\beta}_{(0)} = (1,\ -1,\ -1)^T$, $\boldsymbol{\beta}_{(0)} = (1,\ -1,\ 1)^T$ und $\boldsymbol{\beta}_{(0)} = (1,\ -1,\ 0)^T$.

(l) (3 points) A generalization of the Michaelis-Menten model is the so-called EMAX specification:

$$response_t = \beta_1 + \frac{\beta_2 dose_t^\gamma}{dose_t^\gamma + \beta_3^\gamma}.$$

Estimate this model and plot the regression function in the scatterplot. Test whether the Michaelis-Menten specification is rejected.

(m) (3 points) Two different substances were administered. Now, differentiate between them in the estimation, i.e., estimate:

$$response_t = \beta_1 + \frac{\beta_2 dose1_t}{dose1_t + \beta_3} + \frac{\beta_4 dose2_t}{dose2_t + \beta_5}$$

Generate a graph again. Test using the Wald test whether the two drugs have different effects.[2]

2. Task (0 points) (*Not relevant for the course.*)

3. Task (5 points) NLS: (**Identification in the nonlinear regression model**)

Given is the nonlinear regression model

$$y_t = \beta_1 + \beta_2 x_t^{\beta_3} + u_t, \qquad \mathrm{E}[u_t|x_t] = 0, \qquad x_t > 0, \qquad t = 1,\ldots,n. \tag{5}$$

[2]Hint: This test can be performed using the 'linearHypothesis' function from the 'car' package.

(a) (2 points) Propose a feasible moment vector for estimating the parameters using the method of moments. What about its efficiency?

(b) (2 points) Provide two parameter configurations $\boldsymbol{\beta}_0 = \begin{pmatrix}\beta_{10} & \beta_{20} & \beta_{30}\end{pmatrix}^T$ that are asymptotically non-identifiable.

(c) (1 point) Specify a scenario where identification fails in the sample.

4. Task (8 points) NLS: (**Moment estimation and nonlinear LS method**)

Consider the model

$$y_t = \mathrm{e}^{x_t\beta} + u_t, \qquad \mathrm{E}[u_t|x_t] = 0, \qquad \mathrm{Var}\,(u_t|x_t) = \sigma^2, \qquad t = 1, 2, \ldots, n. \tag{6}$$

(a) (3 points) Determine the optimal instrument w_t for a nonlinear moment estimation in the case of $\beta_0 = 0$.

(b) (2 points) Express the sum of squared residuals as a function $SSR(\beta|\boldsymbol{y}, \boldsymbol{x})$. Compute the first and second derivatives with respect to β.

(c) (3 points) You want to compute the nonlinear LS estimator. Using the Newton-Raphson method for given x_t, y_t, $t = 1, \ldots, n$, and starting value $\beta_{(0)} = 0$, compute the first iteration value $\beta_{(1)}$.

5. Task (0 points) (*Not relevant for the course.*)

6. Task (0 points) (*Not relevant for the course.*)

7. Task (19 points) NLS: **Log(beta)**

Consider the regression model for $\beta > 0$:

$$y_t = \log(\beta)x_t + u_t \qquad \text{mit} \quad \mathrm{E}[u_t|x_t] = 0, \qquad \mathrm{E}[u_t^2|x_t] = \sigma^2. \tag{7}$$

(a) (2 points) Provide $x_t(\beta)$ and $\boldsymbol{X}_t(\beta)$ for this model.

(b) (2 points) Provide the optimal instrument for the nonlinear moment estimator of β. Is the optimal moment estimator applicable here?

(c) (3 points) You want to compute the nonlinear LS estimator. Provide its first-order condition.

(d) (4 points) You start with $\beta_{(0)} = 1$ an iterative Gauss-Newton algorithm. Calculate $\beta_{(1)}$.

(e) (3 points) Calculate the standard error of $\hat{\beta}_{NLS}$ only in terms of the data and $\hat{\beta}_{NLS}$.

You have estimated the model (7) using the nls() function in **R** and obtained the following output.

```
--------------------------------------------------------------
Formula: y ~ log(beta) * x

Parameters:
     Estimate Std. Error t value Pr(>|t|)
beta   1.1841     0.1059   11.18   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
Residual standard error: 0.861 on 99 degrees of freedom

Number of iterations to convergence: 3
Achieved convergence tolerance: 3.775e-09
------------------------------------------------------------
```

It is also known that $\sum_{t=1}^{n} y_t^2 = 76.0378$. You want to test the null hypothesis H_0 : **"$\boldsymbol{x}$ has no effect on $\boldsymbol{y}$".**

(f) (2 points) Perform the t test at the significance level $\alpha = 0.05$.

(g) (3 points) Perform a nonlinear F test for the same null hypothesis.

8. Task (0 points) (*Not relevant for the course.*)

9. Task (0 points) (*Not relevant for the course.*)

10. Task (0 points) (*Not relevant for the course.*)

11. Task (0 points) (*Not relevant for the course.*)

12. Task (0 points) (*Not relevant for the course.*)

13. Task (0 points) (*Not relevant for the course.*)

14. Task (0 points) (*Not relevant for the course.*)

15. Task (14 points) GLS: (**Generalized Least Squares as Moment Estimator**)

Consider the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad \mathrm{E}(\mathbf{u}\mathbf{u}^T|\mathbf{X}) = \boldsymbol{\Omega} \neq \sigma^2\mathbf{I}.$$

The general moment estimator for the matrix $\mathbf{W}$ with $\mathrm{E}(\mathbf{u}\mathbf{u}^T|\mathbf{X},\mathbf{W}) = \boldsymbol{\Omega}$ and $\mathrm{E}(\mathbf{u}|\mathbf{X},\mathbf{W}) = \mathbf{0}$ is determined by

$$\mathbf{W}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \stackrel{!}{=} \mathbf{0}. \tag{8}$$

(a) (2 points) Provide model-based cases where $\mathrm{E}(\mathbf{u}\mathbf{u}^T|\mathbf{X},\mathbf{W}) = \mathrm{E}(\mathbf{u}\mathbf{u}^T)$ holds and where this equation is violated.

(b) (1 point) Show that the moment estimator $\hat{\boldsymbol{\beta}}_{MM} = (\mathbf{W}^T\mathbf{X})^{-1}\mathbf{W}^T\mathbf{y}$ arises from the condition (8).

(c) (3 points) Derive the conditional variance-covariance matrix $\mathrm{Var}(\hat{\boldsymbol{\beta}}_{MM}|\mathbf{X},\mathbf{W})$.

(d) (3 points) Show that the GLS estimator resulting from $\mathbf{W} = \boldsymbol{\Omega}^{-1}\mathbf{X}$ is efficient among general unbiased moment estimators in finite samples.

Assume time series observations $\{y_t, \mathbf{X}_t\}$. Strict exogeneity of $\mathbf{X}$ is now relaxed in favor of $\mathrm{E}(u_t|\mathbf{X}_t) = 0$.

(e) (2 points) What moment condition must a valid ïnstrument"$\mathbf{W}$ fulfill? In which case can this be problematic for the choice $\mathbf{W} = \boldsymbol{\Omega}^{-1}\mathbf{X}$?

(f) (3 points) Show that if the choice $\mathbf{W} = \boldsymbol{\Omega}^{-1}\mathbf{X}$ is permissible, it produces an asymptotically efficient estimator.

16. Task (12 points) GLS: (**Wage Phillips Curve and Autocorrelation in Errors**)

A macroeconomic model of wage setting can be represented as

$$\Delta w_t = \beta_{w0} + \beta_{w1}u_t + \beta_{w2}\Delta a_t + \beta_{w3}\Delta q_t + \beta_{w4}\Delta p_t + \varepsilon_{w,t} \tag{9}$$

where w_t denotes (in natural logarithms) nominal wages, u_t the unemployment rate, a_t labor productivity, q_t producer prices, and p_t the consumer price index. It is assumed that
$\mathrm{E}(\varepsilon_{w,t}|\mathbf{X}) = 0$.

(a) (1 point) Explain the economic intuition behind the equation. Discuss the plausibility of the assumption $\mathrm{E}(\varepsilon_{w,t}|\mathbf{X}) = 0$.

(b) (1 point) In the annual data for Norway `norge_data.csv` , in addition to the variables mentioned above (`d.*` denotes log differences, `*.l1` indicates first lag of the respective variables), there are other potential determinants of wage setting, such as the log wage replacement rate rpr_t, and the wage tax in the industrial sector in levels, $t1_t$, and the changes in daily working hours Δh_t. Read the data into `R` and plot it graphically.

(c) (2 points) Now estimate the parameters of the Phillips curve (9). Plot the residuals as a time series. Examine the residuals graphically for autocorrelation (`acf`) and heteroscedasticity.

(d) (3 points) Calculate standard errors of the model that are consistent even in the presence of heteroscedasticity and autocorrelation. Use the `sandwich` package and the functions `vcovHC` and `vcovHAC`. Calculate the weights for the latter first using `weightsAndrews`:

(i) for the Hansen-White estimator with $p = 3$ (`kernel="Truncated"`, `bw=3`, `prewhite=0`),

(ii) for the Newey-West estimator with $p = 3$ (`kernel="Bartlett"`, `bw=3`, `prewhite=0`), and

(iii) for the Newey-West estimator with automatically calculated bandwidth (no `bw`).

(e) (2 points) Now consider the following model as a possibility:

$$\Delta w_t - \Delta p_{t-1} = \beta_0 + \beta_1 \Delta p_{t-1} + \beta_2 \Delta q_t + \beta_3 \Delta q_{t-1} + \beta_4 u_t + \beta_5 IP_t + \varepsilon_t, \mathrm{E}(\varepsilon_t | \Omega_t) = 0$$

Here, Ω_t denotes the information set at time t (potential explanatory variables up to and including time t, dependent variable up to $t-1$). Interpret the estimated coefficients. Check if there could be misspecification here.

(f) (3 points) Plot the recursive estimators: Estimate the model using the data up to time $t_0 = 1976$. Repeat the estimation with one additional observation each time and plot the coefficient estimators as well as confidence intervals as a function of the end time point. Are there any indications of parameter instability?

17. Task (0 points) (*Not relevant for the course.*)

18. Task (0 points) (*Not relevant for the course.*)

19. Task (0 points) (*Not relevant for the course.*)

20. Task (15 points) GLS: (**Estimation of a Cigarette Demand Function**)

For this task, the panel dataset `Cigarette` from the R package `Ecdat` will be used. You can load it into your workspace using `data(''Cigarette'', package=''Ecdat'')`. The `plm` package provides various panel methods.

(a) (1 point) Load the `Ecdat` package and review the description of the variables in the `Cigarette` dataset (e.g., using `?Cigarette`).

Consider the panel model

$$packpc_{it} = \alpha + \beta_1 avgprs_{it} + \beta_2 \left(\frac{income}{pop}\right)_{it} + u_{it}, \qquad i = 1, \dots, 48, \; t = 1, \dots, 11.$$

Note that here i and t are counted differently from the dataset. The dataset contains observations for the years 1985 to 1995 and for 48 of the 51 sequentially numbered states in the USA. The errors u_{it} can be decomposed into

$$u_{it} = v_i + \varepsilon_{it}.$$

Initially, assume that the unobservable individual-specific effects v_i are random and independent of $avgprs_{it}$ and $\left(\frac{income}{pop}\right)_{it}$.

(b) (2 points) Estimate the model using the random-effects method. Use the `plm` function from the package of the same name.

(c) (3 points) Interpret the parameter estimates and `theta` in the output. Note that `theta` corresponds to λ from the lecture. How could you calculate this from the other information provided in the output?

(d) (2 points) Now assume that the unobserved individual-specific effects v_i are fixed and estimate the model using the fixed-effects method.

(e) (2 points) Interpret the parameter estimates again and compare them with those of the random-effects estimation. Was this result expected from this estimation?

(f) (2 points) Perform the Hausman test using the `phtest` command to test the uncorrelatedness of the individual-specific effects with the other regressors, and interpret the result.

(g) (2 points) Do you see a reason in the present example to choose between the two estimations

(h) (1 point) Do you see any reason, in terms of content/economics, to doubt the exogeneity of the prices $E(u_{it}|avgprs_{it}) = 0$? Do you see a way out?

21. Task (0 points) (*Not relevant for the course.*)

22. Task (0 points) (*Not relevant for the course.*)

23. Task (13 points) IV: (**Simple Instrumental Variable Estimation**)

Let's consider a linear model with k explanatory variables $\mathbf{X}_t$, which do not have to be predetermined:

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t, \quad u_t|\Omega_t \sim IID(0, \sigma^2).$$

You have an $n \times k$ instrumental matrix $\mathbf{W}$ with $\mathbf{W}_t \in \Omega_t$ available.

(a) (2 points) Show that $\hat{\boldsymbol{\beta}}IV := (\mathbf{W}^T\mathbf{X})^{-1}\mathbf{W}^T\mathbf{y} = (\mathbf{X}^T\mathbf{PWX})^{-1}\mathbf{X}^T\mathbf{P_W y}$. Hint: $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ if the inverses exist.

(b) (3 points) The 2SLS estimator $\hat{\boldsymbol{\beta}}_{2sls}$ corresponds to the OLS estimator for $\boldsymbol{\beta}$ in $\mathbf{y} = \hat{\hat{\mathbf{X}}}\boldsymbol{\beta} + \mathbf{u}$. Here, $\hat{\hat{\mathbf{X}}} = (\hat{\hat{\mathbf{x}}}_1, \ldots, \hat{\hat{\mathbf{x}}}_k)$ are the fitted values from the auxiliary regressions $\mathbf{x}_j = \mathbf{W}\boldsymbol{\gamma}j + \mathbf{v}j$, $j = 1, \ldots, k$. Show that $\hat{\boldsymbol{\beta}}IV = \hat{\boldsymbol{\beta}}2sls$ holds.

(c) (3 points) Now let $\hat{\boldsymbol{\delta}}$ be the OLS estimator from $\mathbf{y} = \mathbf{W}\boldsymbol{\delta} + \eta$. The $(k \times k)$ matrix $\widehat{\boldsymbol{\Gamma}} = (\hat{\gamma}_1, , \ldots, , \hat{\gamma}_k)$ contains the OLS coefficient estimators for all k regressions $\mathbf{x}_j = \mathbf{W}\boldsymbol{\gamma}_j + \mathbf{v}j$, $j = 1, \ldots, k$. Show that $\hat{\boldsymbol{\beta}}_{IV} = \widehat{\boldsymbol{\Gamma}}^{-1}\hat{\boldsymbol{\delta}}$.

(d) (2 points) Show that $Q(\hat{\boldsymbol{\beta}}_{IV}) = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{IV})^T\mathbf{P_W}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{IV}) = 0$.

(e) (3 points) Now assume $E[u_t|\mathbf{X}_t] = 0$. Provide the asymptotic covariance matrix of the OLS estimator $\hat{\boldsymbol{\beta}}_{OLS}$ and the IV estimator $\hat{\boldsymbol{\beta}}_{IV}$ under appropriate assumptions. Use the difference of the asymptotic precision matrices to show that $\hat{\boldsymbol{\beta}}_{OLS}$ is at least as efficient as $\hat{\boldsymbol{\beta}}_{IV}$.

24. Task (10 points) IV: (**Generalized IV Estimation**)

You have an $n \times l$ instrument matrix $\mathbf{W} = (\mathbf{W}_1; , \mathbf{W}_2)$, where $\mathbf{W}_1$ is an $n \times l_1$ matrix and $\mathbf{W}_2$ is an $n \times l_2$ matrix ($l = l_1 + l_2$, $l_1 \geq k$).

(a) (2 points) Provide the asymptotic covariance matrix of the generalized IV estimator $\hat{\boldsymbol{\beta}}_{GIV}$ using the entire instrument matrix $\mathbf{W}$. What assumptions do you need for this?

(b) (4 points) Show that $\hat{\boldsymbol{\beta}}_{GIV}$ is asymptotically at least as efficient as the estimator $\tilde{\boldsymbol{\beta}}_{GIV}$, which only uses the instruments $\mathbf{W}_1$.

(c) (4 points) How does the efficiency comparison turn out if for $\mathbf{x}_j = \mathbf{W}_1\boldsymbol{\gamma}_{1j} + \mathbf{W}_2\boldsymbol{\gamma}_{2j} + \mathbf{v}_j$, $\boldsymbol{\gamma}_{2j} = \mathbf{0}$ for all $j = 1, \ldots, k$? Argue asymptotically and in finite samples.

25. Task (11 points) IV: (**Milk demand**)

Consider a linear model of milk demand at the country level

$$q_t = p_t\beta + u_t, \qquad \mathrm{E}(u_t|\Omega_t) = 0, \qquad \mathrm{Var}(u_t|\Omega_t) = \sigma^2, \qquad t = 1, 2, \ldots, n, \tag{10}$$

where q_t represents (logarithmic) milk consumption per capita, p_t denotes (logarithmic) retail price of milk, and β denotes the elasticity of milk demand. A random sample is available.

(a) (1 point) Name two reasons why in the present economic example $\mathrm{E}(u_t|p_t) = 0$ could be violated.

(b) (2 points) Provide a possible instrument. Discuss its suitability.

Now, **one** instrumental variable $z_t \in \Omega_t$ is available. It holds that

$$S_{pz} := \mathrm{E}(p_t z_t), \qquad 0 < |S_{pz}| < \infty,$$
$$S_{pp} := \mathrm{E}(p_t^2), \qquad 0 < S_{pp} < \infty,$$
$$S_{zz} := \mathrm{E}(z_t^2), \qquad 0 < S_{zz} < \infty.$$

(c) (4 points) Under the validity of model assumptions for $n \to \infty$, provide the asymptotic behavior (asymptotic distribution or probability limit, if existent) of the following terms

(i) $\frac{1}{n}\sum_{t=1}^{n} z_t u_t$, (ii) $\frac{1}{n}\sum_{t=1}^{n} z_t p_t$, (iii) $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t u_t$, (iv) $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t p_t$

(d) (2 points) Show (under the validity of necessary assumptions) the consistency of the IV estimator.

Now, you have $l > 1$ instruments $w_{tj} \in \Omega_t$, $j = 1, \ldots, l$ available. The optimal instrument is unknown.

(e) (2 points) In which case is it **not** asymptotically disadvantageous to use only a single instrument w_{ts}, $s \in 1, 2, \ldots, l$ and the simple IV estimator?

26. Task (0 points) (*Not relevant for the course.*)

27. Task (0 points) (*Not relevant for the course.*)

28. Task (47 points) IV: (**Colonization and development**)

(a) (0 points) Read Acemoglu et al. (2001).

(b) (4 points) Briefly explain the relationship between settler mortality and population density during colonial times, institutional arrangements during colonial times, current institutions, and current GDP per capita. Provide a corresponding equation system as well.

(c) (3 points) How is the quality of institutions measured? How is its effect on income economically justified?

(d) (4 points) In Table 2, OLS results of a regression of GDP per capita on institutions and other regressors are provided. Using the equation system formulated in part (b), demonstrate where problems for the OLS estimation could arise.

(e) (4 points) Describe the empirical approach pursued in Table 4. For Table 4, approach (2), provide the regressor matrix $\boldsymbol{X}$ and the instrument matrix $\boldsymbol{W}$. What is the estimator?

(f) (4 points) List necessary assumptions for the validity of the approach and discuss them. How does the article attempt to support these?

(g) (3 points) How do you explain the differences between OLS and IV estimations?

(h) (5 points) Describe the test of overidentification restrictions conducted in Table 8. State assumptions, hypotheses, test statistic, asymptotic distribution.[3] Do you see any issues?

[3] See Davidson & MacKinnon (2004, Section 8.6)

(i) (3 points) Where does the article leave questions unanswered?

Der Datensatz `ajr_data.csv` enthält Originaldaten zu Acemoglu et al. (2001), wovon einige hier aufgelistet sind:

lgdp95	Log Gross Domestic Product per capita PPP 1995
institutions	Measure of institutional quality (see paper)
lsettlermortality	Log settler mortality during colonial times
basesample	Sample used for estimation in Acemoglu et al. (2001)
settlers1900	European settlers in 1900
latitude	'Distance' from the equator (|latitude of the capital|/90)
constrexecut1900	Executive constraints in 1900
constrexec_indep	Executive constraints in the first year of independence
democ1900	Democracy index in 1900
democ_indep	Democracy index in the first year of independence
cont_...	Continent dummies (reference category: America)

(j) (3 points) Load the data `ajr_data.csv` into `R`. Examine it descriptively and graphically. Then, create a dataset containing only the base sample.

(k) (3 points) Run regression (6) from Table 2. Test whether the continent dummies collectively have a significant effect on income.

(l) (5 points) Now, estimate the same specification using the IV method, using settler mortality as an instrument for institutions. Are the continent dummies significant now? Use the `ivreg()` function from the `AER` package.

(m) (4 points) Test the robustness of the parameter estimates by considering additional exogenous control variables, as in Tables 5-7.

(n) (2 points) Now, in further specifications, use the instruments proposed in Table 8. Is there evidence that any of the instruments could be invalid? Test the overidentification restrictions in each case. Evaluate the approach.

29. Task (0 points) (*Not relevant for the course.*)

30. Task (25 points) GMM: (**A GMM calculation example**)

Given is a linear model

$$y_t = x_t\beta + u_t, \qquad \mathrm{E}(u_t|z_t) = 0, \quad t = 1, \ldots, n$$

$$\boldsymbol{S}_{W^T\Omega W} := \lim_{n\to\infty} \mathrm{Var}\Big(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{W}_t^T u_t\Big) = \operatorname{plim} \frac{1}{n}\boldsymbol{W}^T\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{W} = \lim_{n\to\infty}\frac{1}{n}\mathrm{E}(\boldsymbol{W}^T\boldsymbol{\omega}\boldsymbol{W})$$
$$= \lim_{n\to\infty}\frac{1}{n}\mathrm{E}(\boldsymbol{W}^T\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{W}) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\mathrm{E}(\boldsymbol{W}_t^T u_t \boldsymbol{W}_s u_s), \qquad \mathrm{Var}\,(u_t) = \sigma^2. \tag{11}$$

Let $\boldsymbol{W}_t = z_t$ be the only available instrument. You have the following data:

$$\boldsymbol{W} = \boldsymbol{z} = (0.2,\ -0.2,\ 0.3,\ 0.1,\ 0.4,\ 0.8)^T$$
$$\boldsymbol{y} = (0.6,\ -0.8,\ -5.7,\ -2.9,\ -3.4,\ 1.4)^T$$
$$\boldsymbol{x} = (-0.1,\ 0.9,\ -2.3,\ -2.4,\ 0.3,\ 1.1)^T.$$

(a) (3 points) Compute the simple IV estimator for β as well as the residuals of the mode

(b) (2 points) Explain why (11) **does not** imply

$$\boldsymbol{S}_{W^T\Omega W} = \sigma^2\,\mathrm{E}(z_t^2) \tag{12}$$

What additional assumption do you need for this? Calculate an estimator $\widehat{\boldsymbol{S}}_{W^T\Omega W}$ for case (12).

(c) (2 points) Compute $\widehat{\boldsymbol{S}}_{W^T\Omega W}$ assuming a random sample and

$$\mathrm{Var}\,(u_t|z_t) = \sigma_t^2 \neq \sigma^2, \quad t = 1, \ldots, n. \tag{13}$$

(d) (3 points) Compute $\widehat{\boldsymbol{S}}_{W^T\Omega W}$ if the following holds:

$$\mathrm{E}(u_t u_s|z_t, z_s)\begin{cases} \neq 0 & \text{if } |t-s| \leq 1, \\ = 0 & \text{otherwise,}\end{cases} \qquad t, s = 1, \ldots, n$$

How might an estimator for $\boldsymbol{S}_{W^T\Omega W}$ look if all autocovariances are different from zero? (No calculation required in the last case.)

(e) (2 points) Compute an estimator for the variance of $\hat{\beta}_{IV}$ under the random sample assumption and (13).

(f) (4 points) Compute the applicable efficient GMM estimator assuming (13) and instruments $\boldsymbol{W}_t = (z_t, ; z_t^2)$.

(g) (4 points) Test the overidentification restriction in the last case.

(h) (1 point) Under the assumptions made so far, can you construct a fully efficient GMM estimator (GMM with optimal instruments) in each case?

(i) (4 points) Now assume that $\mathrm{E}\, x_t|z_t = e^{z_t}$ and $\mathrm{Var}\,(u_t|z_t) = z_t^2$, $t = 1, \ldots, n$. Determine a transformed model with homoskedastic residuals. What is the optimal instrument for this? Calculate the "fully efficient" GMM estimator.

31. Task (0 points) (*Not relevant for the course.*)

32. Task (0 points) (*Not relevant for the course.*)

33. Task (0 points) (*Not relevant for the course.*)

34. Task (0 points) (*Not relevant for the course.*)

35. Task (0 points) (*Not relevant for the course.*)

36. Task (0 points) (*Not relevant for the course.*)

37. Task (0 points) (*Not relevant for the course.*)

38. Task (38 points) GMM: (**New Keynesian Phillips Curve**)

The New Keynesian Phillips Curve is represented in its linear form as

$$\pi_t = \lambda mc_t + \gamma_f \, \mathrm{E}\, \pi_{t+1}|\Omega_t \tag{14}$$

Here, π_t denotes the inflation rate in period t and mc_t represents the marginal production costs in the economy. λ and γ_f are theoretically positive parameters. For a derivation, see, for example, Romer (2005), and for the subsequent empirical approach, particularly Galí & Gertler (1999).

Now, you are to verify and quantify this relationship using an econometric model, where the labor share ws_t is supposed to approximate the marginal costs (see Galí & Gertler 1999).

$$\pi_t = \lambda ws_t + \gamma_f \pi_{t+1} + u_t \tag{15}$$

The dataset `nkpc_data.csv` contains quarterly US data, including

inf	Quarterly inflation rate of the GDP deflator
ws	Logarithm of the share of labor income in GDP
outp_gap	Quadratically trend-adjusted log GDP
wageinf	Quarterly growth rate of aggregate wage level
lspread	Logarithm of the spread between long-term and short-term interest rates
variable.l...	Lags of the respective variable
variable.p...	Leads of the respective variable.

(a) (4 points) Explain why $\mathrm{E}\, u_t \pi_{t+1} \neq 0$. Why might $mc_t \in \Omega_t$ be violated?

(b) (1 point) Load the data into `R`. Restrict the sample to January 1, 1965, through December 31, 1999.

(c) (1 point) Perform a Generalized Method of Moments (GMM) estimation of the parameters γ_f and λ in (14). Use the first 4 lags of all available variables as instruments.

(d) (3 points) What assumptions must valid instruments satisfy? Do you consider the instruments in the present estimation suitable? (Justification or estimation required!) If necessary, modify the set of instruments.

(e) (2 points) Now, examine the residuals. Are the standard errors and t statistics usable?

(f) (5 points) Now, estimate the same model using the applicable efficient Generalized Method of Moments (GMM). In `R`, you can use the `gmm` function from the package of the same name. Use the Newey-West HAC estimator with 12 lags.[4] Interpret the result.

(g) (3 points) Explain the significance of the so-called `J-statistic`. What does it indicate in this case?

A hybrid version of the Phillips curve, which combines forward-looking with backward-looking price setting, is

$$\pi_t = \lambda mc_t + \gamma_f \operatorname{E} \pi_{t+1}|\Omega_t + \gamma_b \pi_{t-1} \tag{16}$$

Estimate this specification as well.

(h) (4 points) Test whether backward-looking price setting is relevant. How does this test relate to the test for overidentifying restrictions?

(i) (3 points)

(j) (3 points) Attempt GMM estimation with different instrument matrices. Use different time horizons for estimation. Employ other HAC estimators and bandwidths. Is the result robust?

In their original form, the behavioral parameters θ and β (and, for the hybrid Phillips curve, ω) enter the pricing equation nonlinearly. The variables then satisfy, for the New Keynesian Phillips curve (see Galí & Gertler 1999)

$$\operatorname{E} \boldsymbol{W}_t^T (\theta\pi_t - (1-\beta\theta) ws_t - \beta\theta\pi_{t+1}) = 0, \tag{17}$$

while for the hybrid version, the moment conditions are given by

$$\operatorname{E} \boldsymbol{W}_t^T (\phi\pi_t - (1-\omega)(1-\theta)(1-\beta\theta) ws_t - \beta\theta\pi_{t+1}) - \omega\pi_{t-1} = 0 \tag{18}$$

with $\phi := \theta + \omega[1 - \theta(1-\beta)]$.

(k) (3 points) In `R`, create a function `est_fun` that, depending on a parameter vector $(\theta, \beta, \omega)^T$ (argument `theta`) and a dataset `data`, outputs an $(n \times l)$-matrix with the t-th row $\boldsymbol{W}_t(\phi\pi_t - (1-\omega)(1-\theta)(1-\beta\theta) ws_t - \beta\theta\pi_{t+1}) - \omega\pi_{t-1})$.

(l) (3 points) Use this function for a nonlinear GMM estimation of the nonlinear New Keynesian Phillips curve

(m) (3 points) Repeat this for the hybrid version and compare the results.

[4] "The estimator of Newey & West (1987) is a special case of the class of estimators introduced by Andrews (1991). It can be obtained using the "Bartlett"kernel and setting bw to lag + 1."(cf. Help for kernHAC)

39. Task (30 points) ML: (**Poisson Distribution**)

Consider the Poisson-distributed random variables y_t, $t = 1, 2, ..., n$, which are drawn from a random sample. The probability density function of the Poisson distribution is:

$$f_{y_t}(y|\theta) = \begin{cases} \frac{\theta^y}{y!} e^{-\theta} & y = 0, 1, 2, 3, ... \\ 0 & \text{sonst.} \end{cases}$$

The "true" parameter value of the data generating process (DGP) is θ_0. It holds that $\mathrm{E}_0(y_t) = \mathrm{Var}_0(y_t) = \theta_0$, where E_0 and Var_0 denote the expected value and variance under the true DGP.

(a) (3 points) What is the joint probability function $f(y_1, y_2, ..., y_n|\theta_0)$?

(b) (3 points) Determine the log-likelihood function $l(\theta|\boldsymbol{y})$.

(c) (2 points) Calculate the score (or gradient) $g(\theta, \boldsymbol{y}) := \frac{\partial l(\theta|\boldsymbol{y})}{\partial \theta}$.

(d) (2 points) Compute the Hessian matrix (here scalar) $H(\theta) = \frac{\partial^2 l(\theta|\boldsymbol{y})}{\partial \theta^2}$.

(e) (4 points) Show that the following holds:

$$\left. \frac{\partial \, \mathrm{E}_0[l(\theta|\boldsymbol{y})]}{\partial \theta} \right|_{\theta=\theta'} = \mathrm{E}_0[g(\theta', \boldsymbol{y})].$$

(f) (3 points) Show that

$$\theta_0 = \underset{\theta}{\operatorname{argmax}} \; \mathrm{E}_0(l(\theta|\boldsymbol{y})).$$

(g) (2 points) What is the maximum likelihood estimator $\hat{\theta}_{ML}$?

(h) (2 points) In what way can the ML estimator be interpreted as a GMM estimator?

(i) (2 points) How is the variance of the ML estimator calculated?

(j) (2 points) Provide an estimator for the variance of the ML estimator.

(k) (5 points) Provide 3 different methods for estimating the variance of general ML estimators. Determine these for the present model.

40. Task (49 points) ML: (**Exponential Distribution**)

You observe data $\boldsymbol{y} = (0.37, ; 0.13, ; 0.53, ; 0.04, ; 0.14)^T$, which are modeled as a random sample from an exponentially distributed variable. The density is given by

$$f_{y_t}(y|\theta_0) = \begin{cases} \theta_0 e^{-\theta_0 y} & y > 0 \\ 0 & y \leq 0 \end{cases}.$$

The "true" parameter value of the data generating process (DGP) is θ_0. It holds that

$$\mathrm{E}_0(y_t) = \frac{1}{\theta_0}, \qquad \mathrm{Var}_0(y_t) = \frac{1}{\theta_0^2},$$

where E_0 and Var_0 denote the expected value and variance under the true DGP.

(a) (3 points) Calculate the joint density $f_{\boldsymbol{y}}(y_1, y_2, \ldots, y_n|\theta_0)$ as well as the log-likelihood $l(\theta|\boldsymbol{y})$.

(b) (3 points) Is the ML estimator $\hat{\theta}_{ML}$ globally or locally identified for every given sample?

(c) (6 points) Is the parameter θ_0 globally or locally identified?

(d) (3 points) Calculate plim $\frac{1}{n}l(\theta|\boldsymbol{y})$. What statements can be made about the asymptotic identification of θ_0?

(e) (5 points) Provide the Fisher information $\mathbf{I}(\theta)$. Determine the asymptotic Fisher information $\mathcal{I}(\theta)$.

(f) (3 points) Calculate the Hessian matrix (scalar here) $\mathbf{H}(\theta)$ and the asymptotic Hessian matrix $\mathcal{H}(\theta)$.

(g) (1 point) Show that the asymptotic information matrix equality holds.

You want to test the hypothesis using the given dataset:

$$H_0 : \quad \theta = 5 \quad \text{vs.} \quad H_1 : \quad \theta \neq 5$$

(h) (3 points) Calculate the estimator $\hat{\theta}_{ML}$.

(i) (4 points) Provide three possible estimators for the variance of the ML estimator. Choose one.

(j) (5 points) Calculate the Wald test statistic.

(k) (5 points) What is the test statistic of the likelihood ratio test?

(l) (5 points) Determine the Lagrange multiplier test statistic.

(m) (3 points) Make a test decision for each.

41. Task (0 points) (*Not relevant for the course.*)

42. Task (0 points) (*Not relevant for the course.*)

43. Task (0 points) (*Not relevant for the course.*)

44. Task (0 points) (*Not relevant for the course.*)

45. Task (0 points) (*Not relevant for the course.*)

Literatur

Acemoglu, D., Johnson, S. & Robinson, J. A. (2001), 'The colonial origins of comparative development: An empirical investigation', *The American Economic Review* **91**, 1369–1401.

Chakroun, M. (2009), 'Health care expenditure and GDP: An international panel smooth transition approach', *Munich Personal RePEc Archive* **17559**, 1–20.
URL: *http://mpra.ub.uni-muenchen.de/17559/*

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

Galí, J. & Gertler, M. (1999), 'Inflation dynamics: A structural econometric analysis', *Journal of Monetary Economics* **44**, 195–222.

Hansen, B. E. (2012), *Econometrics*.

McMillan, D. G. & Wohar, M. E. (2009), 'Stock return predictability and dividend-price ratio: A nonlinear approach', *International Journal of Finance and Economics* **15**, 351–365.

Nerlove, M. (1963), Returns to scale in electricity supply, *in* C. F. Christ, ed., 'Measurement in Economics - Studies in Mathematical Economics and Econometrics in Memory of Yehua Grunfeld', Stanford Univ. Press, pp. 111–126.

Romer, D. (2005), *Advanced Macroeconomics*, McGraw-Hill.