ECON4160 ECONOMETRICS – MODELLING AND SYSTEMS ESTIMATION

FORTY EXERCISES.

FOR SEMINAR DISCUSSION, INDIVIDUAL TRAINING, EXAM PREPARATION ETC.

Erik Biørn, Department of Economics

Version of August 24, 2011

EXERCISE 1. Specify the regression equation

 $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i,$ $i=1,\ldots,n$,

as a complete econometric model when the x 's are considered as stochastic. Give an interpretation of the assumptions you have specified. Express

$$
m_{yk} = M[y, x_k] = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(x_{ki} - \bar{x}_k) , \qquad k = 1, 2,
$$

by means of β_0 , β_1 , β_2 , the (empirical) variances and the covariances of the x's as well as the covariances between the x's and the u's. Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be the ordinary least squares estimators of β_1 and β_2 . Derive expressions for $\widehat{\beta}_1 - \beta_1$ and $\widehat{\beta}_2 - \beta_2$ and use the result to show that $\hat{\beta}_1$ and $\hat{\beta}_2$ are unbiased and consistent when the standard assumptions of the regression model are satisfied. Which of these assumptions are necessary to ensure that the estimators are

- (a) Gauss-Markov-estimators (MVLUE)?
- (b) unbiased?
- (c) consistent?

EXERCISE 2.

(a) Explain precisely, in relation to a simple regression model with stochastic right hand side variables, what is meant by saying that the right hand side variables are exogenous. If you can give alternative definitions of exogeneity, it would be fine.

(b) Discuss critically the following statement: "In order to use ordinary least squares as a method for estimating a regression model we have to assume that the regression equation is linear and that the disturbances are mutually uncorrelated and have zero expectations and constant variances."

EXERCISE 3. Explain the properties of the ordinary least squares estimators of the coefficients of linear regression models whose disturbances are (i) heteroskedastic, (ii) autocorrelated. Which estimation methods would you recommend in such situations? Give the reason for your answer.

EXERCISE 4. Explain what is meant by multicollinearity. In which ways can multicollinearity frequently be an obstacle in estimating the coefficients of a regression equation? It is often asserted that multicollinearity mainly occurs in analyzing time series data. Do you agree, and if so, why? You may well use examples to illustrate your points. Which characteristics would you use to detect multicollinearity in a specific case?

EXERCISE 5. Consider the regression equation

$$
(i) \t Y = \alpha + \beta X + \gamma Z + u.
$$

In addition we assume that a linear relationship exists between X and Z , of the form

$$
(ii) \t Z = \lambda + \delta X + v,
$$

where u and v are disturbances. We assume that $E(u|X) = E(v|X) = 0$, $E(u^2|X) = \sigma_u^2$, $E(v^2|X) = \sigma_v^2$, $E(uv|X) = \sigma_{uv}$. Show that these assumptions imply that $E(u) = E(v)$ $0, cov(u, X) = cov(v, X) = 0, E(u^2) = \sigma_u^2, E(v^2) = \sigma_v^2, E(uv) = \sigma_{uv}.$ Derive expressions for $cov(u, Z)$ and $cov(v, Z)$. Which of the assumptions with respect to the disturbances above should be satisfied for ordinary least squares estimation of (i) to give consistent estimators of β and γ ?

EXERCISE 6. We are interested in estimating an Engel function for a consumption commodity from data from a survey of households' consumption expenditures. We have, however, only observations in the form of group means for K groups of households (e.g., from a table in a publication). Assume that we formulate our regression equation as follows:

$$
\bar{y}_k = \alpha + \beta \bar{x}_k + u_k, \qquad k = 1, \dots, K,
$$

where \bar{y}_k is the mean expenditure of the relevant commodity in group k and \bar{x}_k is the corresponding mean value of total expenditure. The number of households in group k is n_k $(k = 1, \ldots, K)$, which is assumed to be known. Do you have comments on this model formulation? How would you proceed to estimate the parameters of this Engel function? Try to compare your proposed estimator of β with the one you would have applied if you had had access to the observations from all the $n = \sum n_k$ individual households participating in the survey of households' consumption expenditures.

EXERCISE 7. We want to estimate a (Keynesian) consumption function on the basis of aggregate time series data. The function can be formulated as a linear regression equation on three different ways:

(a)
$$
\frac{C_t}{P_t} = \alpha + \beta \frac{R_t}{P_t} + u_t,
$$

$$
(b) \tC_t = \alpha P_t + \beta R_t + v_t,
$$

(c)
$$
\frac{C_t}{R_t} = \alpha \frac{P_t}{R_t} + \beta + w_t,
$$

where C_t and R_t are, respectively, consumption expenditure and income at current values and P_t is a consumer price index. Which assumptions would you make about the disturbances u_t , v_t and w_t ? Explain how you would estimate the marginal propensity to consume, β , (i) by OLS and (ii) by GLS, in the three cases. What can you say about the properties of the estimators? In the consumption functions above we have assumed that the consumers do not have money illusion, since a proportional change in P_t and R_t is assumed to leave the real value of consumption, C_t/P_t , unchanged. Could you, by modifying the specifications (a), (b) or (c) in suitable ways, use them as a starting point for investigating econometrically whether the consumers have money illusion?

EXERCISE 8. You have been given the task of estimating a linear regression equation between the total capital stock of a production sector and the output of the sector from annual data. You strongly suspects that the disturbances of the equation in any two successive years are positively correlated. Assume that you represent this by means of a first order autoregressive process (AR(1) process).

- (i) Which could be the reasons for having such a suspicion?
- (ii) How would you formulate the regression model?
- (iii) How would you estimate the coefficients of the regression equation?
- (iv) Which are the properties of the OLS in this case?
- (v) Could there be arguments for estimating the model after the variables have been transformed to first-differences?

EXERCISE 9.

(a) Describe one of more methods which you would find useful in investigating whether the disturbances u_t or v_t or w_t in models (a)–(c) in Exercise 7 exhibit heteroskedasticity. (b) Describe one or more methods that you would find useful in order to investigate whether the disturbance of the model you have chosen in Exercise 7, exhibits autocorrelation.

EXERCISE 10. Explain the differences between the Generalized Least Squares (GLS) and the Feasible Generalized Least Squares (FGLS) methods. Explain how you would proceed to apply the latter method in a specific situation. What can you say about the properties of the FGLS method?

EXERCISE 11. Give examples illustrating the use of the GLS in estimating the coefficients of a single regression equation. Can it be convenient to apply this estimation method in situations where two or more regression equations are combined into one equation (SUR)? Explain.

EXERCISE 12. Consider a static model of household consumer demand specified as the linear expenditure systems (LES).

(a) Derive the corresponding cost function and indirect utility function. Which are the relationships between them?

(b) Show that the demand system satisfies Roy's identity.

(c) Some econometricians have taken a parametric specification of the indirect utility function as their point of departure when developing empirical models of consumer demand (f.ex. models based on additive indirect utility functions).

(d) Which arguments may be given for and against such a modelling strategy?

EXERCISE 13. A researcher wants to analyze the substitution between the production factors labour and capital in a sector. He/she chooses for this purpose a CES production function. The model contains, inter alia, the following variables:

 $K =$ Capital input. $L =$ Labour input. $W =$ Wage rate. $C =$ Price of capital services (user cost of capital).

Assume that the producers are price takers and minimize the cost of the two factors for given production. It can be shown that the following relationship holds (derivation is not required):

$$
\ln\left(\frac{K}{L}\right) = a - \sigma \ln\left(\frac{C}{W}\right),\,
$$

where σ is the elasticity of substitution between labour and capital and α is a constant. The researcher wants to analyze the substitution properties econometrically from annual data, but thinks that this static equation is too simple because the factor ratio must be assumed to respond to changes in the price ratio with some sluggishness. He/she therefore will use a dynamic specification, represented by a lag distribution.

The researcher sets $y = \ln(K/L)$ and $x = \ln(C/W)$ and will attempt using three different dynamic versions:

(1)
$$
y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + \varepsilon_t,
$$

where a, b_0, b_1, b_2 are free parameters,

(2)
$$
y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + \varepsilon_t, \text{ where } b_i = b_0 \left(1 - \frac{i}{3} \right), i = 0, 1, 2,
$$

and a and b_0 are free parameters,

(3)
$$
y_t = \alpha + \beta x_t + \lambda y_{t-1} + \varepsilon_t,
$$

where α, β, λ are free parameters and $0 \leq \lambda \leq 1$. [Hint: It may be convenient to consider (3) as a first order linear difference equation in y.

(a) Explain briefly what we, in general, mean by the term lag distribution. Next, explain what sort of lag distributions between x and y which are represented by (1) – (3) .

(b) Give econometric specifications of (1), (2) and (3) and explain how you would estimate (b_0, b_1, b_2) in (1) and (2) and (β, λ) in (3). The researcher has obtained the following estimates of the coefficients in the three equations:

For (1):
$$
(\hat{b}_0, \hat{b}_1, \hat{b}_2) = (-0.40, -0.25, -0.20)
$$
.
For (2): $\hat{b}_0 = -0.45$.
For (3): $\hat{\beta} = -0.21, \ \hat{\lambda} = 0.75$.

Estimate the effect on $\ln(K/L)$ of a change in $\ln(C/W)$ which, according to the three models and results above, are realized

- (i) in the current year,
- (ii) after one year,
- (iii) after two years,
- (iv) after three years,
- (v) in the long run.

Explain briefly what these results say about the substitution between labour and capital in the sector analyzed.

EXERCISE 14. Using a Logit model for qualitative (discrete) response, we want to analyze factors which can determine whether a household uses tobacco or not. Formulate a simple Logit model, and explain briefly its interpretation. Let x_i denote the vector of explanatory variables for household no. i, and let β denote the vector of the corresponding coefficients, which occur in the form $x_i\beta$ in the exponential function in the Logit expression. The explanatory variables include, inter alia, the income, the relative price of tobacco (price of tobacco divided by the consumer price index), the number of persons in four age groups (given below), the age and year of birth of the head person. Estimation based on Norwegian data for 25 180 households observed over the period 1975- 1994 by means of the maximum likelihood method gave the following estimates of the corresponding coefficients in the β -vector and their standard deviations (in parenthesis):

Interpret these results, and explain what they say about the effect of the income, the tobacco price, the number of persons in the four age classes, and the age and year of birth of the head person on the propensity to use tobacco in Norwegian households.

EXERCISE 15. One is frequently interested in analyzing factors which determine qualitative variables. Often such variables are represented by variables which only can take the values zero or one. Give reasons why using a linear regression model may be inconvenient when the endogenous variable is such a binary variable. Describe, preferably by means of an example, a simple Logit model, and explain briefly why it may be more suitable than a classical linear regression model in analyzing the effect of the specified explanatory variables.

EXERCISE 16. Explain precisely the terms structural form and reduced form of a simultaneous, linear equation system. Explain precisely the difference between a simultaneous linear equation system and a system of linear regressions equations.

EXERCISE 17. Consider a market model of a consumer commodity:

- (S) $y_t = \alpha_1 + \beta_1 p_t + \gamma_{11} z_{1t} + u_t,$
- (D) $y_t = \alpha_2 + \beta_2 p_t + \gamma_{22} z_{2t} + \gamma_{23} z_{3t} + v_t,$

where (S) is the supply function, (D) is the demand function, y is the quantity traded, p is the market price, the z 's are three exogenous variables and u and v are disturbances. If you can propose specific interpretations of the z 's it will be fine.

(i) Formulate the market model as a complete econometric model.

(ii) Derive the model's reduced form.

(iii) What are the conditions for the supply function and the demand functions, respectively, to be identifiable?

(iv) A proposal has been made to estimate the model's parameters by first applying OLS on the reduced form equations. Discuss this procedure.

(v) Assume that the quantity supplied is price inelastic. How would you reformulate the model in order to take this into account?

(vi) Can it be convenient to estimate the demand function by means of OLS in a situation as described under (v), and if so, how?

EXERCISE 18. Consider the following simple model of a market for a consumer commodity:

where y_t is the quantity traded of the commodity in year t, p_t is the market price in year t ($t = 1, \ldots, T$), a, b, c, d are unknown constants and where the disturbances u_t and v_t are uncorrelated, have zero expectations and variances equal to σ_u^2 and σ_v^2 , respectively. It is assumed not to be correlation between the disturbances from different years. We want to estimate the price sensitivity of the demand, b , from the T observations of the market price and the traded quantity.

(i) Show that the OLS estimator of the price coefficient in (D), denoted as \hat{b} , has probability limit

$$
\text{plim}(\hat{b}) = \frac{\sigma_v^2 b + \sigma_u^2 d}{\sigma_v^2 + \sigma_u^2}.
$$

(ii) How would you interpret this result? Explain briefly what you understand by simultaneity bias.

A statistician shows the model to three economists, A, B and C, who all assert that it is erroneously specified. They add the following:

A: "In this market not only the commodity price, but also the oil price (which is an important factor price) and the interest rate affects the supply of the commodity."

B: "In this market, the supply is approximately price inelastic, i.e., d is approximately zero."

C: "From my experiences, inter alia based on the situation in other countries, the supply is approximately proportional to the market price, i.e., in (S) $c = 0$ holds as a good approximation."

How would you formulate the model and estimate the coefficient b in order to take into account the comments from, respectively,

```
(iii) economist A, (iv) economist B, (v) economist C?
```
State the properties of the estimators in cases (iii) and (iv). In point (v) a sketch of the argument is sufficient.

(vi) It may be some reason for calling the assumption that u_t and v_t are uncorrelated into doubt. Would any of your conclusions in points (iii) – (v) be changed if you omitted this assumption?

EXERCISE 19. Consider the following partial market model for a commodity:

(1)
$$
p_t = a_0 + a_1 x_t + u_t,
$$

$$
(2) \t x_t = b_0 + b_1 p_{t-1} + v_t,
$$

where p_t and x_t denote the price of and the traded quantity of the commodity in period t and the disturbances u_t and v_t are assumed to be stochastically independent for all t.

(i) Discuss briefly this econometric model.

(ii) Examine whether the structural coefficients are identified.

(iii) Discuss methods for estimating those structural coefficients which you find identifiable.

EXERCISE 20. We want to estimate the marginal propensity to consume of income from aggregate time series data from Norwegian households. Discuss the choice of model, estimation problems and estimation methods for the following cases:

(i) The consumption function belongs to a macro model where both consumption and income are endogenous variables.

(ii) Income is considered as exogenous, but we have to take into consideration that our measurements of this variable contain random errors. Explain in this connection what you understand by random measurement errors.

(iii) Income is considered as exogenous, and both the observations on consumption and income are affected by measurement errors.

EXERCISE 21. We return to the model in Exercise 5. Consider now (i) and (ii) as a simultaneous model with two stochastic equations. What are the conditions for equation (i) being identifiable? Can a connection be said to exist between the multicollinearity problem and the identification problem?

EXERCISE 22. We want to estimate the marginal propensity to consume of income in a consumption function on the basis of cross-section data from a sample of households. We find it permissible to consider the household income as exogenous, but we suspect our income variable to contain a random measurement error.

(a) Compare the following estimation methods:

- (i) regress observed consumption (Y) on observed income (X) ,
- (ii) regress observed income (X) on observed consumption (Y) ,

and examine the asymptotic bias (inconsistency) in the (derived) estimators of the marginal propensity to consume in the two cases.

(b) Assume that you know, from other investigations, that the variance of the measurement error in income, by experience, amounts to 10 per cent of the variance of the observed income. How could you utilize this information to form a consistent estimator of the marginal propensity to consume? Perform the estimation by means of the following data for empirical variances and covariance from the consumption and income surveys of Statistics Norway for 1973:

$$
M_{YY} = 660859,
$$
 $M_{XX} = 399334,$ $M_{YX} = 289837.$

EXERCISE 23. Consider again the market model in Exercise 17.

(a) Explain the indirect least squares (ILS) method in general and how you could use this method to estimate the coefficients of the supply and the demand functions, (S) and (D) .

(b) Another possibility is to use the two-stage least squares method. Explain this method in general and how it can be applied in this specific example. Which conditions should be satisfied for the two-stage least squares method to be applicable?

(c) Explain precisely why an equation which does not satisfy the order condition for identification, cannot be estimated by the two-stage least squares method.

EXERCISE 24. Consider a model with the structural equations

$$
y_{1t} = \alpha_1 + \beta_1 y_{2t} + \gamma_{11} x_{1t} + \gamma_{12} x_{2t} + \varepsilon_{1t},
$$

$$
y_{2t} = \alpha_2 + \beta_2 y_{1t} + \gamma_{13} x_{3t} + \gamma_{24} x_{4t} + \varepsilon_{2t},
$$

where the y 's are endogenous and the x 's are exogenous variables, all of which are observable. Show that this model specification imposes two restrictions on its reduced form coefficients. Describe an estimation method for the model's coefficients which takes these conditions into account.

EXERCISE 25.

(a) Explain briefly, but precisely, the general meaning of the term the identification problem in an econometric context. Explain the order condition, and explain precisely which requirements the model must satisfy in order to use the order condition to investigate whether an equation in a simultaneous linear system is identifiable. What is required for all equations in such a system to be identifiable?

(b) Discuss reasons why it may often be necessary to formulate and use a simultaneous systems of structural equation as the basis for an econometric analysis, even if the primary purpose of the analysis is to estimate the coefficients of one of the structural equations.

EXERCISE 26. It is reasonable to assume, as an approximation, that the production structure in a sector can be described by a Cobb-Douglas production function in labour and capital, with constant returns to scale (scale elasticity equal to 1). Specify econometric models for the following cases:

(a) the producers act as profit maximizers with capital as exogenously determined input, (b) the producers act as cost minimizers with exogenously determined output.

How would you estimate the input elasticity of labour from cross section data from a sample of firms in the two cases? Discuss problems of identification, data problems, and measurement problems that may arise.

EXERCISE 27. Specify an econometric model based on a CES production function in labour and capital

- (a) when the producers are profit maximizers,
- (b) when the producers are cost minimizers with exogenously given output.

Explain how you, in case (b), would (i) estimate the elasticity of substitution between the two factors and (ii) investigate whether the Cobb-Douglas production function is an acceptable simplification of the production structure.

EXERCISE 28.

(a) Explain precisely the meaning of the term instrumental variable in an econometric context.

(b) We want to estimate the marginal propensity to consume of income in a simple keynesian consumption function. The function is part of a simple macro model of an open economy together with, for instance, an investment relation, an import relation and a general budget equation. Specify such a model econometrically. Indicate precisely which variables are exogenous and endogenous. Examine whether the consumption function of your proposed model can be identified. Assume that you have access to mean values and empirical variances and covariances of all the exogenous and endogenous variables in the model. How would you, on the basis of this data set, estimate the marginal propensity to consume by means of a procedure based on instrumental variables? Will all sample means and variances/covariances have to be known for this purpose? What can you say about the properties of the proposed estimators?

EXERCISE 29. Estimating structural coefficients within the framework of econometric models based on cross section data from micro units, may often give results departing widely from those obtained by using models based on aggregated time series data. Discuss, preferably by means of examples, possible explanations of such discrepancies. Describe also ways of combining cross-section data and time-series data in estimating the coefficients of a structural equation.

EXERCISE 30. Assume that you have been given the task to utilize a simultaneous econometric model in formulating predictions of the model's endogenous variables for given values of its exogenous variables. For this purpose it is sufficient to have estimated the model's reduced form. An assertion has been made that in such a situation it is unnecessary to be concerned with estimating the model's structural form and with the problem of identification, since one can formulate the model's reduced form and estimate its coefficients by ordinary least squares directly. Discuss this assertion.

EXERCISE 31. We want to examine how the volume of imports of a good, M , is changed when the real income in the importing country, X , and the ratio of the prices of the import of the good and a related good produced in the importing country, P , is changed. We transform the variables into logarithms and specify the following import function:

$$
(1) \t m_t = \alpha + \beta x_t + \gamma p_t + u_t,
$$

where α , β and γ are unknown coefficients, the subscript t ($t = 1, \ldots, T$), denotes year, $m = \ln(M)$, $x = \ln(X)$, $p = \ln(P)$ and u_t is a disturbance.

(a) Specify a complete econometric model version for (1) when x and p are considered as exogenous and u_t has zero expectation, is serially uncorrelated and

$$
\text{var}(u_t|x_t) = \theta^2 x_t,
$$

where θ is an unknown constant. Explain how you would proceed to estimate β and γ as well as θ in an optimal way. State the reasons for your choice of method.

(b) We are not satisfied with this model description, however. We consider the same import function, (1), but change the specification of the distribution of the disturbances into: u_t has zero expectation,

$$
u_t = \rho u_{t-1} + \varepsilon_t,
$$

where ρ is a known constant, $\rho \in (-1, +1)$. We further assume that ε_t has zero expectation, is serially uncorrelated and

$$
\text{var}(\varepsilon_t | x_t) = \lambda^2 x_t,
$$

where λ is an unknown constant. Give reasons why we may want to make this change. How should we proceed to estimate β and γ as well as λ in an optimal way? Assuming ρ known may seem unreasonable. How would you carry out the estimation if it is unknown?

(c) An analysis of two import functions of the type (1), one for good 1 and one for good 2, shall be performed. We specify them as

(2)
$$
m_{1t} = \alpha_1 + \beta_1 x_t + \gamma_1 p_{1t} + u_{1t},
$$

(3)
$$
m_{2t} = \alpha_2 + \beta_2 x_t + \gamma_2 p_{2t} + u_{2t},
$$

where $\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2$ are unknown coefficients, (m_{1t}, m_{2t}) are the logarithms of the two import volumes and (p_{1t}, p_{2t}) the logarithms of their relative import/home market prices. For simplicity, we now assume that the disturbances, (u_{1t}, u_{2t}) , have zero expectations, constant variances and are serially uncorrelated, but we allow for $cov(u_{1t}, u_{2t}) \neq 0$. State your additional assumptions and explain how you would estimate β_1, β_2 and γ_1, γ_2 in an optimal way.

(d) From the theory leading to (2)–(3) there are reasons to believe that $\beta_1 = \beta_2$, that is, the two import commodities have the same income elasticity. Why would it be unwise not to take this restriction into account? How would you proceed in order to exploit it in your estimation?

(e) Another critique raised against the model in (c) is that the linear functional form is too simple. Choose a model version and explain briefly how you could perform a test to decide whether assuming that the logarithm of the import volumes are linear in the logarithms of the right hand side variables is an acceptable simplification.

EXERCISE 32. The following model for the market for loans from banks to business firms in the USA is specified:

$$
(1) \tQt = \alpha_0 + \alpha_1 RRt + \alpha_2 RSt + \alpha_3 Yt + vt,
$$

$$
(2) \qquad \qquad Q_t = \beta_0 + \beta_1 RR_t + \beta_2 RD_t + \beta_3 X_t + u_t,
$$

where (1) is the supply function for loans, (2) is the demand function for loans, $t(t =$ $1, \ldots, T$ represents the calendar month, u_t and v_t are disturbances and

 $Q_t =$ total commercial loans (in billions of dollars),

 RR_t = average interest rate charged by banks,

 RS_t = Treasury bill rate (considered an alternative rate of returns for banks),

 RD_t = interest rate of bonds issued by firms (considered as the price of alternative financing to firms),

 $X_t =$ a production index for manufacturing (considered an indicator of the activity level of the economy),

 $Y_t =$ total bank deposits (in billions of dollars).

The model's endogenous variables are assumed to be Q_t and RR_t .

(a) Give an interpretation of the model, complete the econometric model description, and examine whether the demand and supply functions are identified. Explain briefly, but precisely what we mean by saying that an equation in an econometric model is identified.

(b) Equations (1) and (2) have been estimated by ordinary least squares (OLS) as well as two-stage least squares (2SLS) from monthly data from the years 1979–1984 ($T = 72$) for the USA. Explain briefly the 2SLS procedure in the present case, and state the most important properties of the two estimation methods. The coefficient estimates are given below, the $\hat{\alpha}_j$, β_j 's denoting OLS estimates and the $\tilde{\alpha}_j$, β_j 's denoting 2SLS estimates (*t*values in parentheses. For 2SLS they are the ratios between the coefficient estimates and their asymptotic standard errors.)

A. For the supply function:

B. For the demand function:

$$
\begin{array}{rcl}\n\widehat{A}_1 &=& -15.99 & (-12.0) & \widehat{A}_1 = & -20.19 & (-12.6) \\
\widehat{A}_2 &=& 2.29 & (5.4) & \widehat{A}_2 = & 2.34 & (5.2) \\
\widehat{B}_3 &=& 36.07 & (14.2) & \widehat{B}_3 = & 46.76 & (14.4) \\
R^2 &=& 0.9768 & (OLS) & R^2 = & 0.9666 & (2SLS)\n\end{array}
$$

Here R^2 denotes the squared coefficient of multiple correlation, computed as the squared coefficient of correlation between Q_t and the computed (predicted) values of Q_t from the respective estimated equations. Do you find the signs of the coefficient estimates reasonable?

(c) Two economists, A and B, who have taken a course in elementary regression analysis, consider the table presenting the above results. Economist A states that "with the data set of 72 observations used, the OLS estimates must be preferable to the 2SLS estimates since the R^2 values of the former exceed those of the latter in the equations". Economist B argues that "we should look at the t-values to decide which of the two methods is the best one; R^2 is not a good quality measure in this case." State whether you agree or disagree with these statements, and give the reasons for your answer. Also indicate whether your would have liked additional calculations to be performed in order to give a better foundation for your conclusions.

(d) As a part of estimation by the 2SLS, the model's reduced form is estimated by OLS. Not infrequently, the R^2 values from this kind of estimation of the reduced form equations are fairly large. Explain briefly why we in such cases often will find that the 2SLS estimates are close to the OLS estimates.

EXERCISE 33. A researcher engaged as an econometric consultant for a firm in a particular industry wants to estimate the input elasticities (marginal elasticities) of labour and capital, α and β , in a Cobb-Douglas production function for the firm. Her data are annual data for T years for the output volume X_t , its price P_t , the labour input L_t , its price W_t , and the input of capital services K_t and its price Q_t from this firm. The production function is specified as:

(1)
$$
X_t = AL_t^{\alpha} K_t^{\beta} e^{v_t}, \quad t = 1, \dots, T,
$$

where A is a constant and v_t is a disturbance. The subscript t denotes year.

Provided that the firm is a price taker and has profit maximization as its objective, a fairly common estimation method for the two input elasticities is the so-called factor-share method or factor-cost-share method. It consists in estimating α and β by, respectively, (proof not required)

(2)
$$
\widehat{\alpha} = \prod_{t=1}^T \left(\frac{W_t L_t}{P_t X_t} \right)^{1/T}, \quad \widehat{\beta} = \prod_{t=1}^T \left(\frac{Q_t K_t}{P_t X_t} \right)^{1/T},
$$

where, for an arbitrary time series Z_t , we let \prod^T $t=1$ $Z_t = Z_1 \cdot Z_2 \cdots Z_T.$

(a) Interpret these estimators. Are $\hat{\alpha}$ and $\hat{\beta}$ unbiased and consistent for α and β , respectively, under the assumptions stated and possible additional assumptions you find it convenient to make? State the reason for your answer. What would be your conclusion if the researcher found that $\hat{\alpha} + \hat{\beta} > 1$?

(b) Now, the firm is not a price taker, but a dominating producer in the market in which it sells its output. We assume that the time series we observe are the results of the firm having behaved as a profit maximizing monopolist and it has been confronted with a demand function with price elasticity d . From this theory it can be shown that the firm will obtain maximal profit by letting

marginal productivity of $L = \frac{W}{R}$ $\frac{W}{P^*}$, marginal productivity of $K = \frac{C}{P^*}$ $\frac{6}{P^*}$

where $P^* = P[1 + (1/d)]$. Assume that the firm's management knows the value of d. Would you still recommend using the factor-share method (2) in estimating α and β ? If not, how would you modify this method to obtain consistent estimators for these parameters when we from our observations find

$$
\prod_{t=1}^{T} \left(\frac{W_t L_t}{P_t X_t} \right)^{1/T} = 0.48, \qquad \prod_{t=1}^{T} \left(\frac{Q_t K_t}{P_t X_t} \right)^{1/T} = 0.20
$$

in the following two cases and point out possible identification problems.:

- (i) when you, from previous studies of the output market, know that $d = -1.8$,
- (ii) when you as an econometrician do not know d.

(c) The observation period is rather long, and it therefore does not seem reasonable to assume, as in (1), that the production technology has been the same in the entire period. We therefore allow for neutral technical progress and replace (1) by

(3)
$$
X_t = A_0 e^{\gamma t} L_t^{\alpha} K_t^{\beta} e^{v_t}, \quad t = 1, ..., T,
$$

where A_0 and γ are positive constants. Would you then modify your conclusions about the estimation procedures for α and β in (a) and (b)? Could you propose a consistent method for estimating γ ?

EXERCISE 34. We want to investigate how the capital stock of two kinds in a production sector, buildings (B) and machinery (M) , depend on the production capacity of the sector. The production capacity is difficult to measure, and we have to prepare ourselves that our observations contain measurement errors. We assume that the measurement errors are random. The following model has been proposed:

(1) $X = X^* + u,$

(2)
$$
K_B = \alpha_B + \beta_B X^* + \varepsilon_B,
$$

(3)
$$
K_M = \alpha_M + \beta_M X^* + \varepsilon_M,
$$

where K_B and K_M are the observed stocks of the two kinds of capital, X is the observed production capacity, X^* is the actual, unobserved magnitude of this variables, u is a random measurement error, and ε_B and ε_M are disturbances. The data are sampled from a cross-section of firms in a certain year. We assume that X^*, u, ε_B and ε_M are mutually uncorrelated and that u, ε_B and ε_M have zero expectations and variances σ_u^2 , σ_B^2 and σ_M^2 , respectively.

(a) Comment briefly on the model, and examine whether $\alpha_B, \beta_B, \alpha_M$ and β_M are identifiable, possibly under which additional assumptions.

(b) Propose estimators of these four parameters.

(c) Assume that there exists a third kind of capital, transport equipment, used by the firms in the sector, with observed value K_T , and assume that we extend the model (1)–(3) by adding the equation

(4)
$$
K_T = \alpha_T + \beta_T X^* + \varepsilon_T.
$$

Would this influence the estimation method you would propose in question (b)?

EXERCISE 35. An econometric version of a very simple keynesian macro model consists of two equations:

$$
(1) \tC_t = \alpha + \beta Y_t + u_t,
$$

$$
(2) \t Y_t = C_t + Z_t.
$$

where C_t, Y_t, Z_t are, respectively, private consumption, GNP, and a variable representing the sum of gross investment, public consumption, and export surplus, all measured at constant prices, and u_t is a disturbance with zero expectation. Subscript t denotes year, and the observations on C, Y and Z cover the years $t = 1, \ldots, T$. All variables are considered as stochastic and we assume

(3)
$$
cov(u_t, Z_s) = 0;
$$

(4)
$$
\text{var}(u_t) = \sigma^2
$$
; $\text{cov}(u_t, u_s) = 0, \quad t \neq s$; $t, s = 1, ..., T$.

We denote the model $(1)–(4)$ as *Model A*.

(a) Interpret assumption (3), and explain briefly why estimating the consumption function (1) by Ordinary Least Squares gives inconsistent estimators of α and β . Explain verbally what we mean by saying that an estimator is inconsistent. Describe a consistent method for estimating the marginal propensity to consume, β and the consumption multiplier of Z , $\Delta C/\Delta Z = \beta/(1-\beta)$, and explain why the method in both cases is consistent.

(b) We are in doubt that the disturbance in Model A satisfies assumptions (4), but we still believe that (3) holds. How would you estimate β if we replace (4) with, respectively,

(5)
$$
\operatorname{var}(u_t|Z_t) = \tau^2 Z_t, \quad \operatorname{cov}(u_t, u_s) = 0, \quad t \neq s,
$$

and

(6)
$$
u_t = \rho u_{t-1} + \varepsilon_t \quad (|\rho| < 1), \quad \text{var}(\varepsilon_t) = \theta^2, \quad \text{cov}(\varepsilon_t, \varepsilon_s) = 0, \ t \neq s,
$$

where τ and θ are positive constants. You may, if you want, assume that ρ is known, but if you could indicate briefly how it could be estimated, it would be fine.

The institution which constructs the national accounts from which the time series C_t , Y_t and Z_t have been taken – let us call it SSB – finds it difficult, from the primary statistics available, based on records from the specific agents in the economy, to construct these time series exactly as the keynesian macro theory prescribes. As alternatives to Model A in order to allow for measurement errors in different ways, three models have been proposed:

Model B: Random measurement errors in all variables:

$$
(7) \tCt* = \alpha + \beta Yt* + ut,
$$

(8)
$$
Y_t^* = C_t^* + Z_t^*.
$$

where C_t^* ^{*}, Y_t^* and Z_t^* are the unobservable theory variables private consumption, GNP, and investment etc. What SSB calculates, and is observed by us are C_t , Y_t and Z_t determined by, respectively,

(9)
$$
C_t = C_t^* + v_{Ct}, Y_t = Y_t^* + v_{Yt}, Z_t = Z_t^* + v_{Zt}.
$$

where v_{Ct} , v_{Yt} and v_{Zt} er random measurement errors.

Model C: Random measurement errors in (C, Y) , systematic error in Z: In this model we assume that (7) and (8) still hold, but that SSB's calculations implies that we observe C_t , Y_t and Z_t determined from, respectively,

(10)
$$
C_t = C_t^* + v_{Ct}, Y_t = Y_t^* + v_{Yt}, Z_t = a_Z + b_Z Z_t^* + v_{Zt},
$$

where a_Z and b_Z are unknown constants and v_{Ct} , v_{Yt} and v_{Zt} are random error terms.

Model **D**: Systematic measurement errors in all variables:

Also in this model we assume that (7) and (8) still hold, whereas SSB's calculations now give us observations on C_t , Y_t and Z_t determined from, respectively,

(11)
$$
C_t = a_C + b_C C_t^* + v_{Ct},
$$

$$
Y_t = a_Y + b_Y Y_t^* + v_{Yt},
$$

$$
Z_t = a_Z + b_Z Z_t^* + v_{Zt},
$$

where a_C, b_C, a_Y, b_Y, a_Z and b_Z are unknown constants and v_{Ct}, v_{Yt} and v_{Zt} are random error terms.

In models **B**, **C** and **D** we assume that u_t , v_{Ct} , v_{Yt} , v_{Zt} and Z_t^* $t_t[*]$ are mutually uncorrelated and that u_t , v_{Ct} , v_{Yt} and v_{Zt} have zero expectations, variances equal to σ_u^2 , σ_{vC}^2 , σ_{vY}^2 and σ_{vZ}^2 , respectively, and non-autocorrelated.

(c) Would it be possible to estimate the marginal propensity to consume of income β consistently from SSB's time series for C_t , Y_t , Z_t in models **B**, **C** and **D**, and if so, how? (d) Going more deeply into SSB's data base you have been able to split the time series for Z_t into I_t = gross investment, G_t = public consumption, and A_t = export surplus, such that $I_t + G_t + A_t = Z_t$. Would you then modify your answer regarding estimation procedures in problem (c), and if so, how?

EXERCISE 36. The following simple model for determination of the quarterly development of wages and prices in a country is specified:

(1)
$$
\dot{p}_t = a_0 + a_1 \dot{w}_t + a_2 \dot{p}_{It} + a_3 \dot{q}_t + \epsilon_t,
$$

(2)
$$
\dot{w}_t = b_0 + b_1 \dot{p}_t + b_2 u_t + b_3 n_t + \delta_t,
$$

where subscript t denotes quarter, the a's and b's are constants, ϵ_t and δ_t disturbances and

 \dot{p}_t = Rate of increase of the consumption price index, pro anno. \dot{w}_t = Rate of increase of the mean wage rate, pro anno. \dot{p}_{It} = Rate of increase of the import price index, pro anno. \dot{q}_t = Rate of increase of the labour productivity, pro anno. u_t = The unemployment rate at the beginning of the quarter. n_t = Share of the labour force unionized at the beginning of the quarter.

In view of the purpose that the model is intended to serve, we consider \dot{p} and \dot{w} as endogenous variables and consider \dot{p}_I , \dot{q} , u and n as exogenous.

(a) Discuss the model briefly, and specify it stochastically. Examine whether the two equations are identifiable.

Two authors presented in an article published in the journal Economica in 1970 estimation results for the price equation (1) and the wage equation (2) based on ordinary least squares and quarterly data for Great Britain for the period 1948:3-1968:2 (80 observations). In parts of the period, the central government had introduced a wage and price control. The authors found, inter alia:

- A. For the price equation (1):
- (i) Using the complete data set:

$$
\begin{array}{rcl}\n\hat{a}_1 &=& 0.562 & (5.53) \\
\hat{a}_2 &=& 0.085 & (4.60) \\
\hat{a}_3 &=& -0.145 & (-3.48) \\
R^2 &=& 0.697 & DW = & 0.946\n\end{array}
$$

(ii) Using only data from quarters where wage and price control was not in effect:

$$
\begin{array}{rcl}\n\widehat{a}_1 &=& 0.851 & (5.52) \\
\widehat{a}_2 &=& 0.073 & (2.93) \\
\widehat{a}_3 &=& -0.092 & (-1.90) \\
R^2 &=& 0.843 & DW = & 1.274\n\end{array}
$$

 $(t$ -values in parenthesis. $DW =$ Durbin-Watson-statistic)

B. For the wage equation (2):

(i) Using the complete data set:

$$
\begin{array}{rcl}\n\bar{b}_1 &=& 0.482 & (5.76) \\
\bar{b}_2 &=& -0.891 & (-1.77) \\
\bar{b}_3 &=& 3.315 & (2.09) \\
R^2 &=& 0.616 & DW = & 0.742\n\end{array}
$$

(ii) Using only data from quarters where wage and price control was not in effect:

$$
\begin{array}{rcl}\n\hat{b}_1 &=& 0.457 & (6.25) \\
\hat{b}_2 &=& -2.372 & (-3.64) \\
\hat{b}_3 &=& 0.136 & (0.07) \\
R^2 &=& 0.856 & DW = & 1.231\n\end{array}
$$

 $(t$ -values in parenthesis. $DW =$ Durbin-Watson-statistic)

(b) Do you find the sign of the coefficient estimates reasonable? Describe briefly how you could perform a test to investigate whether this active price and wage policy had a significant effects on the formation of wages and prices in Great Britain in the actual period, and indicate, without going into details, additional information you would have to possess and supplementary calculations that would be required.

(c) In the article the authors state, inter alia,

(i) "We have used ordinary least squares. In order to be able to use two-stage least squares for estimating (1) and (2) it would be necessary to treat all variables except \dot{p} and \dot{w} as exogenous."

(ii) "If use of two-stage least squares on $(1)-(2)$ should have been possible and we still treated \dot{q} , u and n as endogenous variable, we would have had to specify completely the more comprehensive model to which these two equations belong."

Discuss these two statements.

(d) Assume that you, from the estimation results above or by using some other method that you would prefer, should give the minister of finance in the country an estimate of the effect on the wage increase of an increase in, respectively, the unemployment rate and the rate of increase of the labour productivity. How would you proceed? State the reason for you answer.

EXERCISE 37. We are interested in investigating the relationship between the households' consumption and the gross national product (GNP), i.e., a kind of a macro consumption function, on the basis of time series data for Norway for the period 1865-1939 from the data base of historical data from Statistics Norway. We use the following symbols:

- Y: Gross national product, at constant prices.
- C: Household consumption, at constant prices.

Eight static and three dynamic specifications of the consumption function have been estimated by means of ordinary least squares (OLS):

- (1) $C_t = \alpha + \beta Y_t + u_{1t},$
- (2) $\ln C_t = \delta + \epsilon \ln Y_t + u_{2t},$
- (3) $\Delta C_t = \beta \Delta Y_t + u_{3t},$
- (4) $\Delta C_t = \gamma + \beta \Delta Y_t + u_{4t},$
- (5) $\Delta \ln C_t = \epsilon \Delta \ln Y_t + u_{5t}$
- (6) $\Delta \ln C_t = \phi + \epsilon \Delta \ln Y_t + u_{6t},$
- C_t $\frac{C_t}{Y_t} = \alpha \frac{1}{Y_t}$ (7) $\frac{U_t}{Y_t} = \alpha \frac{1}{Y_t} + \beta + u_{7t},$
- C_t $\frac{C_t}{Y_t} = \alpha \frac{1}{Y_t}$ (8) $\frac{C_t}{Y_t} = \alpha \frac{1}{Y_t} + \beta + \beta' Y_t + u_{8t},$
- (9) $C_t = a + bC_{t-1} + cY_t + u_{9t},$
- (10) $C_t = \alpha + \beta_0 Y_t + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + u_{10t},$
- (11) $C_t = \alpha + \beta_0 Y_t + \beta_1 Y_{t-1} + u_{11t},$

where $\alpha, \beta, \ldots, \phi$ and a, b, c are unknown constants and u_{1t}, \ldots, u_{11t} are disturbances. The result, in the form of edited printouts from the computer programme package Pc-Give, denoted as $EQ(1)-EQ(1)$, is given at the end of the set of questions. The estimate of the standard deviation of the disturbance is denoted as "sigma".

(a) Explain briefly the relationship between the disturbances u_{1t}, u_{3t}, u_{7t} and between the disturbances u_{2t}, u_{5t} , and which properties u_{1t} would have to possess if OLS estimation based on (7) should give Gauss-Markov-estimators of α and β . Which interpretation would you give of the intercept term γ in (4) and the intercept term ϕ in (6)?

(b) Over such a long time span as 75 years, containing, inter alia, periods with good and bad business prospects and with a varying degree of uncertainty in the economy, it is not unlikely that the disturbance variance in (1) has varied. Explain how you would proceed to investigate this and which additional calculations such an investigation would require.

(c) From the estimation results for the consumption functions (1) – (6) it can be argued that autocorrelation in the disturbances is a markedly more prevalent phenomenon when the consumption function is estimated in level form than when it is transformed to first differences before OLS estimation is performed. Do you agree, and how would you explain this? Would you from this recommend estimation on difference form rather than on level form?

(d) Would you, from the estimation results for the consumption functions (1) – (6) , support or reject (i) a hypothesis that the level of the household consumption has been an approximately constant share of GNP and (ii) a hypothesis that the increase in consumption has been a constant share of the increase in GNP over the long observation period? Perform a corresponding investigation for public consumption, denoted as G_t , on the basis of the results reported as $EQ(1^*)-EQ(6^*)$ after Exercise 38.

(e) Do the estimation results under $EQ(9)-EQ(11)$ indicate that there is a sluggishness in the adjustment of the consumption to the income? State the reason for your answer, and explain what is meant my the terms short-run and long-run propensity to consume of income. Explain what kind of lag-distribution (9) describes. Estimate these two parameters from the results under $EQ(9)-EQ(11)$ after Exercise 38. Are the estimators you propose, consistent?

EXERCISE 38 (continuation of EXERCISE 37). The data set also contains:

- G: Public consumption, at constant prices.
- I: Investment, at constant prices.
- X: Export, at constant prices.
- M: Import, at constant prices.

We now consider G, I, X as exogenous variables.

(a) In equation $EQ(12)$ at the end the consumption function (1) is estimated by twostage least squares, with G and I used as instruments for GNP. In equation EQ(13), (4) is estimated by two-stage least squares, with ΔG and ΔI as instruments for the increase in GNP. Explain this procedure and give your opinion on whether it is sensible.

(b) In equation EQ(14) at the end an attempt is made to estimate, by OLS, an import function in the form of first differences (i.e. with ΔM as left hand side variable and ΔC , ΔG and ΔI as right hand side variables). In equation EQ(15), the same function is estimated by two-stage least squares, with ΔC considered as endogenous and ΔG and ΔI as exogenous. In the last case, ΔG , ΔI , and ΔX are used as instruments. Explain the procedure in the last case, and give your conclusions from the estimation results about the three marginal propensities to import.

EQ(1) Modelling c by OLS The estimation sample is: 1865 to 1939 Coefficient Std.Error t-value t-prob Part.R^2 Constant 6885.14 646.0 10.7 0.000 0.6088 y 0.685457 0.01159 59.1 0.000 0.9795 sigma 2659.99 RSS 516516181 R^2 0.979544 F(1,73) = 3496 [0.000]** log-likelihood -696.863 DW = 0.228 no. of observations 75 no. of parameters 2
mean(c) 40482.8 var(c) $3.36669e+008$ $mean(c)$ 40482.8 $var(c)$ --

EQ(2) Modelling lnc by OLS The estimation sample is: 1865 to 1939

Coefficient Std.Error t-value t-prob Part.R^2 Constant 1.06799 0.1214 8.80 0.000 0.5147 lny 0.884891 0.01136 77.9 0.000 0.9881 sigma 0.0496391 RSS 0.179875153 R^2 0.988103 F(1,73) = 6063 [0.000]** log-likelihood 119.816 DW = 0.228 no. of observations 75 no. of parameters 2 mean(lnc) 10.5079 var(lnc) 0.201593 --- EQ(3) Modelling dc by OLS The estimation sample is: 1866 to 1939 Coefficient Std.Error t-value t-prob Part.R^2 dy 0.719745 0.04998 14.4 0.000 0.7396 sigma 1265.1 RSS 116835137 log-likelihood -633.073 DW = 2.32 no. of observations 74 no. of parameters 1 mean(dc) 922.176 var(dc) 5.21326e+006 -- EQ(4) Modelling dc by OLS The estimation sample is: 1866 to 1939 Coefficient Std.Error t-value t-prob Part.R^2 Constant -72.7657 166.8 -0.436 0.664 0.0026 dy 0.731182 0.05669 12.9 0.000 0.6979 sigma 1272.18 RSS 116527118 R^2 0.697945 F(1,72) = 166.4 [0.000]** log-likelihood -632.976 DW = 2.31 no. of observations 74 no. of parameters 2 mean(dc) 922.176 var(dc) 5.21326e+006 -- EQ(5) Modelling dlnc by OLS The estimation sample is: 1866 to 1939 Coefficient Std.Error t-value t-prob Part.R^2 dlny 0.951038 0.06180 15.4 0.000 0.7644 sigma 0.0235392 RSS 0.0404489229 log-likelihood 172.934 DW = 2.23 no. of observations 74 no. of parameters 1 mean(dlnc) 0.021408 var(dlnc) 0.00186157 -- EQ(6) Modelling dlnc by OLS The estimation sample is: 1866 to 1939 Coefficient Std.Error t-value t-prob-Part.R^2
Constant -0.00235126 0.003281 -0.717 0.476 0.0071
0.980110 0.07410 13.2 0.000 0.7085 sigma 0.023618 RSS 0.0401624424 R^2 0.708452 F(1,72) = 175 [0.000]** log-likelihood 173.197 DW = 2.23 no. of observations 74 no. of parameters 2 mean(dlnc) 0.021408 var(dlnc) 0.00186157 -- EQ(7) Modelling c/y by OLS The estimation sample is: 1865 to 1939 Coefficient Std.Error t-value t-prob Part.R^2
Constant 0.769203 0.01380 55.7 0.000 0.9770
1/y 3261.26 480.4 6.79 0.000 0.3870 sigma 0.0492623 RSS 0.177154859 R^2 0.386971 F(1,73) = 46.08 [0.000]** log-likelihood 120.388 DW = 0.169 no. of observations 75 no. of parameters 2 mean(cdy) 0.85456 var(cdy) 0.0038531 -- EQ(8) Modelling c/y by OLS The estimation sample is: 1865 to 1939 Coefficient Std.Error t-value t-prob Part.R^2
Constant 1.14272 0.02657 43.0 0.000 0.9625
1/y -4140.27 563.2 -7.35 0.000 0.4288

y -3.66814e-006 2.517e-007 -14.6 0.000 0.7468 sigma 0.0249588 RSS 0.0448517325 R^2 0.844794 F(2,72) = 196 [0.000]** log-likelihood 171.9 DW = 0.72 no. of observations 75 mean(cdy) 0.85456 var(cdy) 0.0038531 -- EQ(9) Modelling c by OLS The estimation sample is: 1866 to 1939 Coefficient Std.Error t-value t-prob Part.R^2

Constant 2570.59 647.2 3.97 0.000 0.1818

0.601241 0.06423 9.36 0.000 0.5524

y 0.288427 0.04295 6.71 0.000 0.3884 sigma 1787.69 RSS 226904926 R^2 0.990821 F(2,71) = 3832 [0.000]** log-likelihood -657.633 DW = 1.78 no. of observations 74 no. of parameters 3 mean(c) 40791.9 var(c) 3.34053e+008 -- EQ(10) Modelling c by OLS The estimation sample is: 1868 to 1939 Constant Coefficient Std.Error t-value t-prob Part.R⁻²

Constant 6627.58 712.5 9.30 0.000 0.5636

y_1 0.0697683 0.1689 0.413 0.681 0.0025

y_2 -0.159472 0.1688 -0.945 0.348 0.0131

y_3 0.350236 0.1487 2.35 0.021 0.0764 sigma 2573.55 RSS 443750169 R^2 0.981237 F(4,67) = 876 [0.000]** log-likelihood -664.991 DW = 0.329 no. of observations 72 no. of parameters 5 mean(c) 41425.6 var(c) 3.28475e+008 --- EQ(11) Modelling c by OLS The estimation sample is: 1868 to 1939 Coefficient Std.Error t-value t-prob Part.R^2
Constant 7223.78 683.9 10.6 0.000 0.6179
y 0.616561 0.1284 4.80 0.000 0.2504
y_1 0.0665599 0.1343 0.496 0.622 0.0035 sigma 2646.38 RSS 483230117 R^2 0.979568 F(2,69) = 1654 [0.000]** log-likelihood -668.06 DW = 0.264 no. of observations 72 no. of parameters 3 mean(c) 41425.6 var(c) 3.28475e+008 -- EQ(1*) Modelling g by OLS The estimation sample is: 1865 to 1939 Coefficient Std.Error t-value t-prob-Part.R^2
Constant -1583.25 198.6 -7.97 0.000 0.4653
0.139897 0.003565 39.2 0.000 0.9547 sigma 818.012 RSS 48847498 R^2 0.954733 F(1,73) = 1540 [0.000]** log-likelihood -608.423 DW = 0.507 no. of observations 75 no. of parameters 2 mean(g) 5273.79 var(g) 1.4388e+007 -- EQ(2*) Modelling lng by OLS The estimation sample is: 1865 to 1939 Coefficient Std.Error t-value t-prob-Part.R^2
Constant -7.32874 0.3653 -20.1 0.000 0.8465
1.46491 0.03420 42.8 0.000 0.9617 sigma 0.149388 RSS 1.62912252 R^2 0.961733 F(1,73) = 1835 [0.000]** log-likelihood 37.1839 DW = 0.32 no. of observations 75 no. of parameters 2 mean(lng) 8.2987 var(lng) 0.56763 ---

EQ(3*) Modelling dg by OLS

The estimation sample is: 1866 to 1939 Coefficient Std.Error t-value t-prob Part.R^2 dy 0.0494036 0.02042 2.42 0.018 0.0742 sigma 516.926 RSS 19506545.8 log-likelihood -566.843 DW = 1.35 no. of observations 74 no. of parameters 1 mean(dg) 193.716 var(dg) 247206 --- EQ(4*) Modelling dg by OLS The estimation sample is: 1866 to 1939 Coefficient Std.Error t-value t-prob–Part.R^2
Constant 160.904 65.56 2.45 0.017 0.0772
dy 0.0241134 0.02228 1.08 0.283 0.0160 sigma 500.006 RSS 18000423.1 R^2 0.016008 F(1,72) = 1.171 [0.283] log-likelihood -563.87 DW = 1.36 no. of observations 74 no. of parameters 2 mean(dg) 193.716 var(dg) 247206 --- EQ(5*) Modelling dlng by OLS The estimation sample is: 1866 to 1939 Coefficient Std.Error t-value t-prob Part.R^2 dlny 0.422283 0.1851 2.28 0.025 0.0665 sigma 0.0705215 RSS 0.363049645 log-likelihood 91.7379 DW = 1.77 no. of observations 74 no. of parameters 1 mean(dlng) 0.0336292 var(dlng) 0.00412476 -- EQ(6*) Modelling dlng by OLS The estimation sample is: 1866 to 1939 Coefficient Std.Error t-value t-prob Part.R^2 Constant 0.0334052 0.009045 3.69 0.000 0.1593 dlny 0.00924111 0.2043 0.0452 0.964 0.0000 sigma 0.0651093 RSS 0.305223916 R^2 2.84243e-005
F(1,72) = 0.002047 [0.964] log-likelihood 98.1572 DW = 1.98
no. of observations 74 no. of parameters 2
mean(dlng) 0.0336292 var(dlng) 0.00412476
----------------------------EQ(12) Modelling c by 2SLS The estimation sample is: 1865 to 1939 Instruments: g, i Coefficient Std.Error t-value t-prob Constant 6511.29 651.7 9.99 0.000 y Y 0.693084 0.01172 59.2 0.000 sigma 2667.87 RSS 519578581 Reduced form sigma 2103.6 no. of observations 75 no. of parameters 2 no. endogenous variables 2 no. of instruments 3 mean(c) 40482.8 var(c) 3.36669e+008 -- EQ(13) Modelling dc by 2SLS The estimation sample is: 1866 to 1939 Instruments: dg, di Coefficient Std.Error t-value t-prob Constant -258.170 185.3 -1.39 0.168 dy Y 0.867436 0.07610 11.4 0.000 sigma 1322.23 RSS 125877046 Reduced form sigma 1492.7 no. of observations 74 no. of parameters 2 no. endogenous variables 2 no. of instruments 3 mean(dc) 922.176 var(dc) 5.21326e+006 --- EQ(14) Modelling dm by OLS The estimation sample is: 1866 to 1939

23

Coefficient Std.Error t-value t-prob Part.R^2
-444 811 132 6 -3.35 0.001 0.1385 Constant -444.811 132.6 -3.35 0.001 0.1385 dc 0.805837 0.07938 10.2 0.000 0.5955
dg 0.0783570 0.2657 0.295 0.769 0.0012 dg 0.0783570 0.2657 0.295 0.769 0.0012 0.231318 sigma 998.364 RSS 69771216.5 R^2 0.825003 F(3,70) = 110 [0.000]** log-likelihood -613.998 DW = 2.57 no. of observations 74 no. of parameters 4 mean(dm) 417.081 var(dm) 5.38783e+006 -- EQ(15) Modelling dm by 2SLS The estimation sample is: 1866 to 1939 Instruments: dg, di, dx Coefficient Std.Error t-value t-prob Constant -393.067 159.3 -2.47 0.016 dc Y 0.391022 0.2613 1.50 0.139 dg 0.687570 0.4759 1.44 0.153 di 0.706447 0.027 sigma 1177.11 RSS 96991653.6 Reduced form sigma 1555.6
no. of observations 74 no. of parameters 4 74 no. of parameters 4
2 no. of instruments 4 no. endogenous variables mean(dm) 417.081 var(dm) 5.38783e+006

EXERCISE 39. We are interesting in analyzing the capital-labour substitution in two manufacturing industries (2-digit SIC industries) from aggregate input data from the USA for the years 1958-1996 ($T = 39$ observations). The input quantities are in millions of dollars at constant 1992 prices, the price indexes are normalized to $1992 = 1$, and subscript t denotes year. At the end of the set of questions are a slightly edited printout from PcGive, and a comprehensive table of critical values for the Durbin-Watson test (DW). The variables which occur in the econometric models to be considered, are $[ln = the natural logarithm]:$

 $y_{1t} = \ln\left(\frac{K_{1t}}{I}\right)$ L_{1t} $\overline{ }$ = log of (capital input quantity/labour input quantity) in Industry 1, $y_{2t} = \ln\left(\frac{K_{2t}^{10}}{L}\right)$ L_{2t} $\left\langle \right\rangle$ $=$ log of (capital input quantity/labour input quantity) in Industry 2, $x_{1t} = \ln\left(\frac{P_{K1t}^{2e}}{P}\right)$ P_{L1t} $\overline{ }$ = log of (index for capital price/index for labour price) in Industry 1, $x_{2t} = \ln \left(\frac{P_{K2t}^{L1t}}{P} \right)$ P_{L2t} ╲ = log of (index for capital price/index for labour price) in Industry 2,

We assume that the typical firms in the two industries have a production technology of the CES (Constant Elasticity of Substitution) form and minimizes input costs for given output. It can then be shown from the optimizing conditions that

$$
\begin{split} \ln\biggl(\frac{K_{1t}}{L_{1t}}\biggr) &= \alpha_1 + \beta_1\ln\biggl(\frac{P_{K1t}}{P_{L1t}}\biggr) + u_{1t}, \\ \ln\biggl(\frac{K_{2t}}{L_{2t}}\biggr) &= \alpha_2 + \beta_2\ln\biggl(\frac{P_{K2t}}{P_{L2t}}\biggr) + u_{2t}, \end{split}
$$

where $(-\beta_1)$ and $(-\beta_2)$ are the elasticity of substitution between capital and labour in the two industries, and u_{1t} and u_{2t} are disturbances. (Proof not required.)

Using the simplified notation, the system of factor input equations can be written as

(12)
$$
y_{1t} = \alpha_1 + \beta_1 x_{1t} + u_{1t},
$$

(13)
$$
y_{2t} = \alpha_2 + \beta_2 x_{2t} + u_{2t}, \qquad t = 1, ..., T.
$$

(a) Consider the two (logarithmic) price ratios x_{1t} and x_{2t} as exogenous and assume that u_{1t} and u_{2t} are correlated. OLS (Ordinary Least Squares) estimates of Equations (1) and (2) are given in PRINTOUT, PART A, at the end of Exercise 40. Corresponding results when the two equations are estimated as a system of regression equations by FGLS (Feasible Generalized Least Squares) are shown in PRINTOUT, PART B, at the end of Exercise 40. Interpret the two sets of results and explain why the estimates differ.

(b) The standard errors of the coefficient estimates in PRINTOUT, PART B are lower than those in PRINTOUT, PART A. Do you find this an expected result? There is not a similar improvement in the t-values. Do you find the latter finding surprising? Explain briefly.

(c) Are there signs that the assumption of non-autocorrelated disturbances in Equations (1) and (2) is violated? Explain briefly.

(d) The Cobb-Douglas production function is the special case of the CES function where the elasticity of substitution is one $(-\beta_1 = -\beta_2 = 1)$. Would you reject the Cobb-Douglas hypothesis from the results in the printouts? Maybe you would need more detailed output to properly answer this question? State briefly the reason for your answer.

An extension of Equations $(1)-(2)$, where both relative input prices are assumed to enter both input equations, is also considered. The model is then specified as

(14)
$$
y_{1t} = \alpha_1 + \beta_{11}x_{1t} + \beta_{12}x_{2t} + v_{1t},
$$

(15)
$$
y_{2t} = \alpha_2 + \beta_{21}x_{1t} + \beta_{22}x_{2t} + v_{2t}, \qquad t = 1, ..., T.
$$

where v_{1t} and v_{2t} are disturbances.

(e) OLS estimates of Equations (3) and (4) are reported in PRINTOUT, PART C of the printout. Corresponding results when the two equations are estimated jointly as a system of regression equations by FGLS are shown in PRINTOUT, PART D, at the end of Exercise 40. Compare these two sets of results with those obtained in PRINTOUT, PART A and PRINTOUT, PART B and state your conclusion.

(f) A colleague examining PRINTOUT, PART A, . . . ,PRINTOUT, PART D claims that since Industry 1 and Industry 2 produce outputs which are strongly different, the restriction $cov(u_{1t}, u_{2t}) = 0$ should have been imposed on Equations (1)–(2) and $cov(v_{1t}, v_{2t}) = 0$ should have been imposed on Equations (3)–(4). Would your conclusions above then have been different? Do you agree with you colleague that a difference in the nature of the outputs from the two industries is a valid reason for imposing these zero covariance restrictions? Explain briefly your argument.

EXERCISE 40 (continuation of EXERCISE 39). Assume now that an assumed feedback from the rest of the economy, say from the way the labour market functions, gives reason to believe that the input price ratio in Industry 2 affects the input price ratio in Industry 1, but that there is no feedback the other way. To account for this we extend the equation system (1)–(2) by adding a third equation, making x_{1t} endogenous, so that the modified system becomes

$$
y_{1t} = \alpha_1 + \beta_1 x_{1t} + u_{1t},
$$

\n
$$
y_{2t} = \alpha_2 + \beta_2 x_{2t} + u_{2t},
$$

\n
$$
x_{1t} = \gamma + \delta x_{2t} + \varepsilon_t,
$$

where ε_t is a disturbance.

(a) Give a complete specification of this model. PRINTOUT, PART E at the end gives estimation results for the first $[EQ(1)]$ and third equation $[EQ(3)]$ of the latter model when, because of the assumed endogeneity of x_{1t} , we use x_{2t} as instrument for x_{1t} when estimating the first equation. Explain briefly why x_{2t} is a valid instrument and why we do not need to reestimate the second equation in the model?

(b) Are there situations in which, when the three-equation model is the appropriate specification, you would prefer the methods used in PRINTOUT, PART A and PRINT-OUT, PART B to those used in PRINTOUT, PART E? Explain briefly.

(c) Consider the corresponding three-equation system obtained by making x_{1t} in Equations (3) – (4) endogenous:

$$
y_{1t} = \alpha_1 + \beta_{11}x_{1t} + \beta_{12}x_{2t} + v_{1t},
$$

\n
$$
y_{2t} = \alpha_2 + \beta_{21}x_{1t} + \beta_{22}x_{2t} + v_{2t},
$$

\n
$$
x_{1t} = \gamma + \delta x_{2t} + \varepsilon_t,
$$

where you assume that $cov(v_{1t}, v_{2t}) \neq 0$. Is it possible to estimate the coefficients in the two first equations consistently? Does you answer depend on whether $cov(v_{1t}, \varepsilon_t)$ and $cov(v_{2t}, \varepsilon_t)$ are zero or not?

(d) To account for a possible delayed response of the factor input ratio to changes in the factor price ratio, the following extension of Equation (1) in Problem 1 has been proposed:

$$
y_{1t} = \alpha_1 + \beta_1 x_{1t} + \lambda_1 y_{1,t-1} + u_{1t}, \qquad |\lambda_1| < 1,
$$

PRINTOUT, PART F of the printout gives OLS estimates for this equation. Can you from this printout (i) estimate the short-run and long-run elasticity of substitution between labour and capital in Industry 1 consistently, and (ii) test whether the long-run elasticity significantly exceeds the short-run elasticity? Which of the two elasticity estimates is closest to the estimate obtained from the corresponding static equation in PRINTOUT, PART A? Could you explain your finding?

(e) At the end of the exercise is shown – for various combinations of the number of observations (T) and the number of coefficients in the equation (K) – the lower (dL) and the upper (dU) bounds of the 5% critical values of the Durbin-Watson test $(K < T-4)$. Give, supported by this table and what you know about the relationship between residuals and disturbances, a brief intuitive explanation of why the difference between the upper and lower bound $(dU - dL)$ would have been smaller if you had had twice as long time series, i.e., $T = 78$ rather than $T = 39$, with the same value of K. What do you think will happen to $(dU - dL)$ when T goes to infinity for a fixed K?

PRINTOUT, PART A EQ(1) Modelling ln(K1/L1) by OLS. The estimation sample is: 1958 to 1996 Coefficient Std.Error t-value t-prob Part.R^2 Constant -1.64376 0.03979 -41.3 0.000 0.9788 ln(PK1/PL1) -0.814122 0.05918 -13.8 0.000 0.8365 sigma 0.148029 RSS 0.810768809 R² 0.836457 F(1,37) = 189.2 [0.000]
20.1914 DW 0.535 log-likelihood 20.1914 DW 100.1 (0.535
no. of observations 39 no. of parameters 2 39 no. of parameters EQ(2) Modelling ln(K2/L2) by OLS. The estimation sample is: 1958 to 1996 Coefficient Std.Error t-value t-prob Part.R^2 Constant -0.720182 0.03993 -18.0 0.000 0.8980 ln(PK2/PL2) -0.859562 0.07035 -12.2 0.000 0.8014 sigma 0.155971 RSS 0.900100587 0.801366 F(1,37) = 149.3 [0.000]
18.1532 DW 0.312 log-likelihood 18.1532 DW 0.312
no. of observations 39 no. of parameters 2 no. of parameters
******************** ** PRINTOUT, PART B System of regression equations, estimated by FGLS. Version 1. The estimation sample is: 1958 to 1996 (1) Equation for: ln(K1/L1) Coefficient Std.Error t-value t-prob Constant -1.654590 0.03975 -41.6 0.000
1n(PK1/PL1) -0.794065 0.05904 -13.4 0.000 ln(PK1/PL1) -0.794065 0.05904 -13.4 0.000 sigma = 0.148259 (2) Equation for: ln(K2/L2) Coefficient Std.Error t-value t-prob Constant -0.734906 0.03991 -18.4 0.000
1n(PK2/PL2) -0.826284 0.07029 -11.8 0.000 -0.826284 $signa = 0.156442$ Correlation between residuals in equations for ln(K1/L1) and ln(K2/L2): 0.14657 ** PRINTOUT, PART C EQ(1) Modelling ln(K1/L1) by OLS. The estimation sample is: 1958 to 1996 Coefficient Std.Error t-value t-prob Part.R^2 Constant -1.61147 0.02380 -67.7 0.000 0.9922
1n(PK1/PL1) -0.420474 0.05855 -7.18 0.000 0.5889 ln(PK1/PL1) -0.420474 0.05855 -7.18 0.000 0.5889 $ln(PK2/PI.2)$ sigma 0.0873797 RSS 0.274867547 R¹ 0.944555 F(2,36) = 306.6 [0.000]
41.2844 DW 0.662 log -likelihood 41.2844
no. of observations 39 $\overline{}$ no. of parameters

EQ(2) Modelling ln(K2/L2) by OLS. The estimation sample is: 1958 to 1996

System of regression equations, estimated by FGLS. Version 2. The estimation sample is: 1958 to 1996

(1) Equation for: ln(K1/L1) Coefficient Std.Error t-value t-prob
Constant -1.611470 0.02380 -67.7 0.000
1n(PK1/PL1) -0.420474 0.05855 -7.18 0.000
1n(PK2/PL2) -0.553437 0.06606 -8.38 0.000 sigma = 0.0873797

Correlation between residuals in equations for ln(K1/L1) and ln(K2/L2): 0.88602 Correlation between actual and fitted values of ln(K1/L1): 0.97188 Correlation between actual and fitted values of ln(K2/L2): 0.92623 **

PRINTOUT, PART E

EQ(1) Modelling ln(K1/L1) by IVE. The estimation sample is: 1958 to 1996

DESCRIPTIVE STATISTICS. The sample is 1958 to 1996 (39 obs.)

no. of observations 38 no. of parameters 3 ***

Durbin-Watson 5 % Critical Values (dL=lower,dU=upper):

 $T = NO$. OF OBS. $K = NO$. OF COEF. (INCL. INTERCEPT)

