
Applied Econometrics

Prof. Dr. Joachim Grammig

University of Tübingen

summer term 2012

1. Introduction

What is econometrics?

...there are several aspects of the quantitative approach to economics, and no single one of these aspects, taken by itself, should be confounded with econometrics. Thus, econometrics is by no means the same as economic statistics. Nor is it identical with what we call general economic theory, although a considerable portion of this theory has a definitely quantitative character. Nor should econometrics be taken as synonymous with the application of mathematics to economics. Experience has shown that each of these three view-points, that of statistics, economic theory, and mathematics, is a necessary, but not by itself a sufficient condition for a real understanding of the quantitative relations in modern economic life. It is the unification of all three that is powerful. And it is this unification that constitutes econometrics.

Ragnar Frisch, *Econometrica*, (1933), 1, pp.1-2

What is econometrics?

Econometrics = economic statistics \cap economic theory \cap mathematics

Probabilistic approach, not deterministic

Conceptual:

- Data perceived as realizations of random variables
- Parameters are real numbers, not random variables
- Joint distributions of random variables depend on parameters

7 justifications for linear regression

(Angrist and Pischke, 2008, Ch. 1/2)

- Structural model
- Population regression
- Linear conditional expectation function (CEF)
- Smallest mean squared error (MSE) approximation to a nonlinear CEF
- Smallest MSE prediction of dependent variable using a linear forecast function
- (Rubin) causal model

Regression analysis “structural model”

dependent variable = constant + $\beta \times$ key regressor
+ $\gamma' \times$ control variables
+ unobservable component/ residual

Equation derived from economic theory.

Equation derived from a theoretical model

① Parameters β have a structural interpretation, e.g.:

- return on education
- cost efficiency
- marginal propensity to consume
- efficiency of marketing expenditures
- population effect

② Hypothesis testing e.g.:

- return on education $< r$ (market interest rate)
- marginal propensity to consume = 1
- efficiency of marketing expenditures = 0

What's behind the variables?

dependent variable	key regressor	control variables	unobserved component
<u>human capital theory:</u>			
income individual i	years of schooling	experience, age	unobservable abilities
<u>theory of production:</u>			
cost of production utility company i	produced 1000kWh company i		efficiency of the organization
<u>macro theory:</u>			
consumption year t	disposable income year t	interest rate	unobservable characteristics year t (interest rate outlook)

3 examples

Example 1: Glosten-Harris model

- Evolution of financial asset prices
(*Journal of Financial Economics*, 1988, 21 (1), pp.123-142)
- Public and private information
- Notation:
 - Transaction price: P_i
 - Indicator of transaction type: $Q_i = \begin{cases} 1 & \text{buyer initiated trade} \\ -1 & \text{seller initiated trade} \end{cases}$
 - Trade volume: v_i
 - Drift parameter: μ
 - Earnings/costs of the market maker: c
 - Unobserved component (public information): ε_i

Glosten-Harris model (2)

- Efficient price:

$$m_i = \mu + m_{i-1} + \varepsilon_i + Q_i z_i, \quad z_i = z_0 + z_1 v_i$$

- Private information: $Q_i z_i$
- Public information: ε_i
- Trading volume: v_i
- Market maker sets:

$$\text{Sell price (ask): } P_i^a = \mu + m_{i-1} + \varepsilon_i + z_i + c$$

$$\text{Buy price (bid): } P_i^b = \mu + m_{i-1} + \varepsilon_i - z_i - c$$

⇒ Transaction price change

$$\Delta P_i = \mu + z_0 Q_i + z_1 v_i Q_i + c \Delta Q_i + \varepsilon_i$$

Goal: Estimation of structural parameters $\theta = (\mu, z_0, z_1, c)'$

Example 2: Mincer equation

$$\ln(WAGE_i) = \beta_1 + \beta_2 S_i + \beta_3 TENURE_i + \beta_4 EXPR_i + \varepsilon_i$$

- Notation:
- Logarithm of the wage rate: $\ln(WAGE_i)$
- Years of schooling: S_i
- Experience in the current job: $TENURE_i$
- Experience in the labor market: $EXPR_i$

⇒ β_2 : return to schooling

Linear factor asset pricing models

Asset pricing theory \Rightarrow

$$\mathbb{E} \left(R_{t+1}^{ej} \right) = \beta^j \times \lambda$$

$R_{t+1}^{ej} = R_{t+1}^j - R_{t+1}^f$: expected excess return of asset j
 $\beta^j = (\beta_1^j, \dots, \beta_K^j)$: exposure of asset j to factor k risk
 $\lambda = (\lambda_1, \dots, \lambda_K)$: something proportional to price of factor k risk

$$R_{t+1}^j = \frac{x_{t+1}^j}{p_t^j} = \frac{p_{t+1}^j + d_{t+1}^j}{p_t^j}$$

Linear factor asset pricing models

$$\mathbb{E} \left(R_{t+1}^{ej} \right) = \beta^j \times \lambda$$

with single risk factor f (e.g. CAPM)

$$\beta^j = \frac{\text{Cov} (R^{ej}, f)}{\text{Var}(f)}$$

K risk factors: $\beta^j = \mathbb{E} (ff')^{-1} \mathbb{E} (R^{ej} f)$

CAPM and Fama-French model

CAPM $f = R^{em} = R^m - R^f$ “excess return of market portfolio”

Fama-French model $f = (R^{em}, HML, SMB)'$

f contains excess returns $\Rightarrow \lambda = [\mathbb{E}(f_1), \dots, \mathbb{E}(f_K)]'$

$\Rightarrow \mathbb{E}(R_{t+1}^{ej}) = \beta^j \mathbb{E}(R_{t+1}^{em})$ CAPM

$\mathbb{E}(R_{t+1}^{ej}) = \beta_1^j \mathbb{E}(R_{t+1}^{em}) + \beta_2^j \mathbb{E}(SMB_{t+1}) + \beta_3^j \mathbb{E}(HML_{t+1})$ FF

“Compatible regression” (for Fama-French model)

$$\begin{aligned} R_{t+1}^{ej} &= \beta_1^j R_{t+1}^{em} + \beta_2^j SMB_{t+1} + \beta_3^j HML_{t+1} + \varepsilon_{t+1} \\ &= (\beta_1^j, \beta_2^j, \beta_3^j) \times \begin{bmatrix} \mathbb{E}(R_{t+1}^{em}) \\ \mathbb{E}(HML_{t+1}) \\ \mathbb{E}(SMB_{t+1}) \end{bmatrix} + \varepsilon_{t+1} \end{aligned}$$

$$\mathbb{E}(\varepsilon_{t+1} | R_{t+1}^{em}, SMB_{t+1}, HML_{t+1}) = 0 \Rightarrow \mathbb{E}(\varepsilon_{t+1}) = 0 \quad \text{by LTE}$$

no constant/ intercept in regression equation implied by theory \Rightarrow

$$\mathbb{E}(R_{t+1}^{ej}) = \beta_1^j \mathbb{E}(R_{t+1}^{em}) + \beta_2^j \mathbb{E}(SMB_{t+1}) + \beta_3^j \mathbb{E}(HML_{t+1})$$

Justification B: Population regression

Y : dependent variable

(e.g. wage)

X : random vector ($K \times 1$)
of explanatory variables
(regressors)

(gender, age, union, ex-
perience)

f_{YX} joint density in population

Y_i, X_i : i^{th} draw from population

$$f_{YX} = f_{Y_i X_i} \quad \forall i$$

Justification B: Population regression

Population regression coefficients (PRC) from

$$\operatorname{argmin}_{\{\check{\beta}\}} \mathbb{E} \left(Y_i - X_i' \check{\beta} \right)^2$$

PRC $\tilde{\beta}$ solve F.O.C. $\mathbb{E} \left[X_i \left(\underbrace{Y_i - X_i' \tilde{\beta}}_{\tilde{\varepsilon}_i} \right) \right] = 0$ • [Note]

$$\Rightarrow \tilde{\beta} = \mathbb{E}(X_i X_i')^{-1} \mathbb{E}(X_i Y_i)$$

$$Y_i = X_i' \tilde{\beta} + \tilde{\varepsilon}_i$$

- $\tilde{\varepsilon}_i$:
- population regression residual
 - constructed $\tilde{\varepsilon}_i = Y_i - X_i' \tilde{\beta}$, “no life of its own”

Interpretation of $\tilde{\beta}$?

(AP notation: $\check{\beta} = b$, $\tilde{\beta} = \beta$, $\tilde{\varepsilon}_i = e_i$)

For one constant and single regressor

$$X_i = \begin{pmatrix} 1 \\ X_{i2} \end{pmatrix}$$

$$\tilde{\beta} = \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{pmatrix}$$

$$\Rightarrow \tilde{\beta} = \mathbb{E}(X_i X_i')^{-1} \mathbb{E}(X_i Y_i) = \begin{pmatrix} \mathbb{E}(Y_i) - \tilde{\beta}_2 \mathbb{E}(X_{i2}) \\ \frac{\text{Cov}(Y_i, X_{i2})}{\text{Var}(X_{i2})} \end{pmatrix}$$

Population regression $\hat{=}$ linear projection

PRC $\hat{=}$ projection coefficients

“Regression anatomy” formula (Frisch-Waugh)

For $X_i = (1, X_{i2}, \dots, X_{ik}, \dots, X_{iK})'$

$$\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2, \dots, \tilde{\beta}_k, \dots, \tilde{\beta}_K)'$$

$$\tilde{\beta}_k = \frac{\text{Cov}(Y_i, \tilde{X}_{ik})}{\text{Var}(\tilde{X}_{ik})} \quad (\text{bivariate regression})$$

\tilde{X}_{ik} : residual from population regression of X_{ik} (dependent variable) on $X_{i.k}$ (including constant)

$X_{i.k}$: X_i without X_{ik}

$$X_{ik} = \tilde{\beta}'_{i.k} X_{i.k} + \tilde{X}_{ik} \quad \tilde{\beta}_{i.k} = \mathbb{E}(X_{i.k} X'_{i.k})^{-1} \mathbb{E}(X_{i.k} Y_i)$$

$\tilde{\beta}_{i.k}$: bivariate slope coefficient for k^{th} regressor after “partialling out” all other regressors.

• [Note]