Instructor: **Anna Houštecká / Yaroslav Korobka** Updated: June 20, 2025
Course: **Econometrics II**

1 Panel data models

Problem 1

Suppose that the random effects model $y_{it} = x_{it}'\beta + \eta_i + v_{it}$ is to be estimated with a panel in which the groups have different numbers of observations. Let T_i be the number of observations in group i. Show that the pooled least squares estimator is unbiased and consistent despite this complication.

Problem 2

Consider $y_{it} = x_{it}'\beta + \eta_i + v_{it}$, $i = 1, \dots, N$, $t = 1, \dots, T$, where $v_{it} \sim \mathcal{N}(0, \sigma^2)$ and $\beta = 0$. Write out the likelihood for estimating η_i and σ^2, and show that the MLE estimator $\hat{\sigma}^2$ is biased when $T < \infty$.

Problem 3

Consider $y_{it} = \mathbb{1}\{x_{it}\beta + \eta_i + v_{it} \geq 0\}$, where the errors v_{it} have the logisic cdf. Consider $T = 2$, $x_{i1} = 0$ and $x_{i2} = 1$, and show that the sufficient statistic for η_i is $y_{i1} + y_{i2} = 1$, i.e. conditioning on $y_{i1} + y_{i2} = 1$ implies that the MLE does not depend on η_i.

Problem 4

Derive the bias of the OLS estimator for α in a dynamic panel of the form $y_{it} = \alpha y_{it-1} + \eta_i + v_{it}$. Are there any conditions on α that should hold for the estimator to be well-defined?

Problem 5

Consider the panel AR(1) model with individual effects,

$$y_{it} = \alpha y_{it-1} + \eta_i + v_{it}$$

where $\eta_i \sim \text{i.i.d.}(0, \sigma^2)$ and $v_{it} \sim \text{i.i.d.}(0, \sigma^2)$ are mutually independent, and for all i we have $y_{i0} = 0$. Derive $\text{var}\left[y_{it}\right]$ for $t = 1, \dots, T$.

Problem 6

Assume that we are in the AR(1) dynamic model setup such that

$$y_{it} = \alpha y_{it-1} + \eta_i + v_{it}$$

but now our v_{it} follows an MA(1) process such that

$$v_{it} = w_{it} + b w_{it-1},$$

where $w_{it} \sim \text{i.i.d.}(0, \sigma_w^2)$ (i.e. v_{it} is serially correlated). Show that in this case the instrument y_{it-2} is not a valid instrument for estimating α with GMM in first differences, while the instruments y_{it-j} for $j \geq 3$ remain valid.

Problem 7

We have data for a panel of companies on gross investment expenditures I_{it} and net capital stock K_{it}. We model the investment rate $y_{it} = I_{it}/K_{it}$ as

$$\left(\frac{I_{it}}{K_{it}}\right) = \alpha \left(\frac{I_{it-1}}{K_{it-1}}\right) + \eta_i + v_{it},$$

and Table 1 shows the results of estimating the model in *levels* by OLS and WG, and the model in *first differences* with one instrument, two instruments, and all Arellano-Bond instruments. For the last two estimators, it also shows the Sargan test statistic and the m_2 statistic for second-order serial correlation in the residuals from the estimated model.

Table 1: Estimation results (703 firms, 4966 observations)

	OLS	WG	2SLS DIF	GMM DIF	GMM DIF
	(1)	(2)	(3)	(4)	(5)
$\hat{\alpha}$	0.2669	-0.0094	0.1626	0.1593	0.1560
	(.0185)	(.0181)	(.0362)	(.0327)	(.0318)
m_2				0.52	0.46
Sargan test				0.36	0.43
Instruments			$(I/K)_{t-2}$	$(I/K)_{t-2}$	$(I/K)_{t-2}$
				$(I/K)_{t-3}$	$(I/K)_{t-3}$
					$\vdots$
					$(I/K)_1$

a) For each of the models in columns (2) and (3), write down the estimated equation(s).

b) Comment on the estimates of α in each of the columns. Are the results in line with theory (in terms of possible bias of the different estimators)? Why do we need to use instruments?

c) Comment on the standard errors of the last three estimators. Are the results in line with theory?

d) For the two GMM estimators (column (4) and (5)), what do you conclude from the two specification tests? What are these tests' null hypotheses and why are these useful to run?

Problem 8

Formulate a linear dynamic panel regression with a single weakly exogenous regressor, and AR(2) feedback in place of AR(1) feedback (i.e. when two most recent lags of the left side variable are present at the right side). Describe the algorithm of estimation of this model.

2 Causal inference