Problem Set Applied Economics
Panel Data

1. We use the file fatality.gdt in `gretl`. It contains data for the US for 48 states: the unit is a state (variable *state*), and the period is a year (variable *year*).

   We will use only years 1982 and 1988. We will use the following simple model with individual heterogeneity:

   $$MR_{it} = \beta_0 + \beta_1 beertax_{it} + \delta_0 d88_t + a_i + u_{it},\ t = 1982, 1988 \qquad (1)$$

   where MR is computed multiplying by 10.000 the variable *mrall* in order to represent the mortality rate (deaths per 10000 individuals), *beertax* is the tax on a case of beer in 1988 dollars and $d88$ is a dummy variable for the year 1988.

   i) Explain which factors are captured by $d88_t$ and the unobserved effects a_i.

   ii) Estimate the model (1) using OLS pooling both cross-sections. Interpret the results. Are the assumptions needed for OLS consistency reasonable in this case? What bias would you expect if those assumptions are not verified?

   iii) Estimate the model with individual heterogeneity using a first differences estimator. Start from model (1), obtain the model in differences, and estimate the model. Interpret the results. Would it be possible to estimate the effect of the population in each state with this method?

   iv) Estimate the model with individual heterogeneity using the within estimator. Start from model (1), obtain the model in deviations with respect to the mean, and estimate the model. Interpret the results.

   v) Estimate the model with individual heterogeneity considering a_i as a dummy variable for state i. Define the corresponding dummy variables and estimate the model. Interpret the results.

   vi) Compare the three estimations for β_1 obtained using the panel dimension.

   vii) A change in taxes may reflect changes in economic conditions. Omitting this type of variables may generate a bias in OLS estimations. Include as an additional variable in model (1) the unemployment rate (*unrate*). Estimate the model using the within estimator, interpret the results and comment the differences in the estimated effect with the one obtained in the simple model.

2. (Based on Stock and Watson) We use the file **Guns.gdt**. It is a balanced panel of data on 50 US states, plus the District of Columbia (for a total of 51 "states"), by year for 1977 - 1999. Each observation is a given state in a given year. We use it to study the effect of the "shall-carry" laws on violent crimes. These laws instruct local authorities

to issue a concealed weapons permit to all applicants who are citizens (and some other restrictions).

The outcome variable is *vio*: violent crime rate (incidents per 100,000 individuals), and the treatment dummy is *shall*: dummy variable that takes the value one if the state has a shall-carry law in effect

(a) Estimate a pooled OLS regression of $ln(vio)$ on *shall*, and another regression adding the variables *incarc_rate*, *density*, *avginc*, and *pop*.

    i. Interpret the coefficient on the variable *shall* in the first regression.

    ii. Does the estimated effect change when you add other regressors?

    iii. Mention an example of a variable that differs between states but it is constant (or almost) in the period, and therefore could be causing an omitted variable bias in the estimations.

(b) Starting for the most complete model in the previous question, obtain an estimator of the effect of the law including individual fixed effects and time fixed effects. Does the estimated effect change? Are the time effects jointly significant?

(c) Interpret your results. What conclusions would you draw about the effects of concealed-weapons laws on the crime rates?

3. In this exercise we are going to use the data set wagepan.gdt from Vella and Verbeek (1998) to review the estimate of the trade union effect. A primary goal of trade unions is to maintain and improve workers' terms and conditions through collective bargaining with employers. Standard models suggest that the union's success in creating a wage premium depends, in large part, on their bargaining strength, and the overall union premium in a given economy may not be significant.

In the US, at a workplace where a majority of workers have voted for union representation, a committee of employers and union representatives negotiate a contract and individual negotiation is prohibited. Once the workers' committee and management have agreed on a contract, it is then put to a vote of all workers at the workplace. If approved, the contract is usually in force for a fixed term of years, and when that term is up, it is then renegotiated between employees and management.

The data we are going to use comprise 545 men who worked in every year from 1980 through 1987 in the United States. Some variables in the data set change over time: experience, marital status, and union status are the three important ones. Other variables do not change: race and education. Consider the unobserved effects model

$$lwage_{it} = \theta_t D_t + \beta x_{it} + \pi z_i + a_i + u_{it} \tag{2}$$

where θ_t simply denotes different year intercepts, $D_t = \mathbf{1}(year = t)$, and a_i is the unobserved worker effect. Moreover, x_{it} includes:

$$
\begin{aligned}
exper \quad &= \text{labor market experience} \\
expersq \quad &= exper^2 \\
married \quad &= 1 \text{ if married} \\
union \quad &= 1 \text{ if in trade union}
\end{aligned}
$$

and z_i includes

$$\begin{aligned} black &= 1 \text{ if black} \\ hisp &= 1 \text{ if hispanic} \\ educ &= \text{years of schooling} \end{aligned}$$

The data set also includes nr, which refers to the person's identifier, $year$ (goes from 1980 to 1987), and 1-digit occupational code $occup$.

(a) Explain which effects parameters θ_t and unobserved effects a_i are likely to capture.

(b) If unions are successful in their wage negotiations with employers, what should be the sign of β_{union}?

(c) Estimate the equation by pooled OLS. Do you find any evidence for a union effect? Are the assumptions required for these estimates to be consistent plausible? If not, what would be the asymptotic bias you would expect in the union estimate?

(d) Now estimate the model in first differences. Can we estimate the returns to education in FD? Why? What about race effects and experience?

(e) Comment what the FD results of the union effect suggest on the correlation between $union$ and a_i.

(f) Test whether the residuals of the FD model are serially correlated by regressing the following model:
$$\Delta u_{it} = \rho \Delta u_{i,t-1} + \epsilon_{it}$$

In view of these results, what do you think about the efficiency of FD?

(g) Estimate the following model (FE):

$$lwage_{it} = \theta_t D_t + \beta x_{it} + \gamma_i D_i + u_{it} \tag{3}$$

where D_i is a dummy variable for worker i. Explain which effects are captured by the (fixed) parameters γ_i. Comment the results for β_{union}.

(h) Carry out part (f) for the FE model. What do you think about the efficiency of FE in this case?

(i) Estimate a RE model. Comment.