***Question One:***

**This problem employs a dataset on labor markets in 23 OECD countries for the years 1980 to 1998.**

The variables used in the analysis (followed by descriptive statistics) are:

1. Productivity index [*prod*] = An index measuring country i’s economic output (GDP) per hour worked in year t, normalized such that each country’s index = 100 in 1995.

2. Unemployment rate [*unr*] = The total number of unemployed workers in country i and year t divided by the total number of labor force participants in that country and year, multiplied by 100.

3. Union density [*ud*] = The ratio of total reported union members (minus retired and unemployed members) in country i and year t to the total number of employees earning wages or salaries in that country and year, multiplied by 100.

4. Public sector growth [*gempl*] = The one-year percentage growth (from year t-1 to year t) in public sector employment in country i (measured as a proportion, 0 to 1).

5. USD exchange rate growth [*usd*]: The one-year percentage growth (from year t-1 to year t) in the value of country i’s currency relative to the US dollar (measured as a proportion, 0 to 1).

6. Labor force (1K) [*lf*]: The total number of labor force participants in country i and year t, in thousands.



Table 1 below presents results from six regressions.

**Table 1: Fixed Effects Estimates**

**Dependent Variable: Productivity index**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|   | (1) | (2) | (3) | (4) | (5) | (6) |
| Unemployment rate | 1.543\*\*\* | -0.0310 | 1.543\*\*\* | -0.0310 | 1.543\*\*\* | -0.0310 |
|  | (0.227) | (0.137) | (0.227) | (0.137) | (0.446) | (0.299) |
| Union density | -0.822\*\*\* | 0.0523 | -0.822\*\*\* | 0.0523 | -0.822\*\*\* | 0.0523 |
|  | (0.0813) | (0.0515) | (0.0813) | (0.0515) | (0.218) | (0.118) |
| Public sector growth | -115.7\*\*\* | 23.28 | -115.7\*\*\* | 23.28 | -115.7\* | 23.28 |
|  | (40.39) | (26.58) | (40.39) | (26.58) | (59.88) | (18.64) |
| USD exchange rate change | 14.24\* | 2.039 | 14.24\* | 2.039 | 14.24\* | 2.039 |
|  | (8.361) | (6.642) | (8.361) | (6.642) | (8.115) | (5.629) |
| Labor force (1K) | 0.000764\*\*\* | -0.000591\*\*\* | 0.000764\*\*\* | -0.000591\*\*\* | 0.000764\* | -0.000591 |
|  | (0.000111) | (0.000104) | (0.000111) | (0.000104) | (0.000375) | (0.000350) |
| Constant | 3.437 | 152.1\*\*\* | 99.54\*\*\* | 85.00\*\*\* | 99.54\*\*\* | 85.00\*\*\* |
|   | (14.70) | (13.24) | (5.327) | (3.046) | (14.36) | (7.799) |
| Observations | 361 | 361 | 361 | 361 | 361 | 361 |
| Estimator | OLS | OLS | AREG | AREG | XTREG | XTREG |
| Time Effects | No | Yes | No | Yes | No | Yes |
| R-squared | 0.476 | 0.907 | 0.476 | 0.907 | 0.397 | 0.893 |
| Robust standard errors in parentheses |  |  |  |  |  |  |
| \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 |  |  |  |  |  |  |

1. While there are no missing years in the dataset, there are missing observations for some of the variables.
2. If there were no missing values for any variables, how many observations (country-years) would there be for every variable in the summary table of descriptive statistics presented above?
3. Given the number of observations for each variable shown in the summary table, knowing there are no missing years in the data, and knowing that Stata regression drops a case when there are any missing values for any variable for a given country in a given year, what is the maximum number of countries that can be used in a regression analysis (assuming nothing is done to replace missing values)?
4. Given that there are 19 years of data in the regression analyses presented in Table 1, how many countries were used in the analyses?
5. Could we estimate the effect of *usd* on *prod* with FE if the value of every country’s currency (relative to the US dollar) remained the same over the sample time period? Why or why not? Please answer in 2-3 sentences.
6. Write the general equations for the specifications in columns (1) and (2). Use lowercase **b** for the regression coefficients and, where appropriate, **a** to indicate fixed effects and/or **T** to indicate time effects. Use the variable names presented in brackets [ ] on the prior page, and use subscripts as appropriate. You do not have to include an error term.

Column 1:

Column 2:

1. Using the models estimated without time effects, interpret:

A. The effect of a 5-percentage-point increase in union density.

B. The effect of a 10-percent increase in the growth of the public sector.

1. Compare the specifications with time effects to those without time effects. What do the differences in the statistically significant coefficients imply about the time effects? Note that the time effects are jointly statistically significant with a p-value of 0.00. Please answer in 2-3 sentences.

***Question Two:***

Table 2 presents results of a study[[1]](#footnote-1) of the effect of differences in the fraction of new immigrants on crime rates in U.S. metropolitan area (MA’s) over nine years.
 ^

* 1. Write the general equation for the regression in column 2. Use β for the regression coefficients (*not* the actual numbers in column 2), use the variable names presented in the table in brackets [ ], and use subscripts as appropriate. If appropriate, use MA and T as fixed effects.
	2. Using the results in column 2:
		1. Ignoring significance, what is the effect of a twenty (20 percentage point or .2 fraction) increase in new immigrants on the overall crime rate?
		2. Form a 95% confidence interval around the effect you’ve just calculated.
	3. Two parts:
		1. In what *two* ways does the coefficient on Fraction of new immigrants [IMM] differ between columns 1 and 2?
		2. *Why* does it differ and *what* does this indicate about the estimates in columns 1 and 2?
	4. Using the results in column 2: For an MA with 10% (.10 fraction) Hispanic population, what is the effect of a one percentage point (.01 fraction) increase in the percent female on the metropolitan crime rate?
	5. Using the results in column 2, what is the effect of a one percent increase in population of an MA on the overall crime rate of the MA?
	6. Two parts:
		1. What hypotheses do the p values for F’s in column 2 at the bottom of the table test? (Hint: There are two different p values (F’s) and thus two different hypotheses.)
		2. What do you conclude from the tests?

|  |  |  |
| --- | --- | --- |
| Table 2 |  |  |
| Regression coefficients: Log metropolitan area (MA) overall crime rate (CR) on various variables |
|  | Dependent Variable: Ln MA overall crime rate |
|  | (1) | (2) |
| *Fraction* New Immigrants [(IMM)]z | 13.23 | -1.066 |
|  | (1.973) | (2.017) |
| Mean Age [(AGE)] |  | -0.011 |
|  |  | (0.010) |
| Fraction Female [(F)] |  | -1.853 |
|  |  | (0.723) |
| Fraction Hispanic [(H)] |  | 2.174 |
|  |  | (0.310) |
| (Fraction Female) \*(Fraction Hispanic) [(FH)] |  | 2.100 |
|  |  | (0.532) |
| Log Population [(Ln POP)] |  | 0.072 |
|  |  | (0.023) |
| Log mean Wage [(LnW)] |  | -0.663 |
|  |  | (0.156) |
| Year Dummies (p-Value)~ | 0.06 | 0.028 |
|  |  |  |
| MA Fixed effects (p Value)~ | <0.04 | < 0.01 |
|  |  |  |
| R-Square | 0.1515 | 0.4972 |
|  |  |  |
| Observations | 339 | 339 |
| Source: Calculations from Current Population Surveys (CPSs) and Uniform Crime Reports (UCR) |
| Notes: Robust Standard errors are parentheses and constant included but not shown. |  |  |
| ~ p-value from an F-test |  |  |
| z: "*Fraction*" varies from 0 to 1 and differs in measurement from percent, which varies from 1 to 100. |

1. Adopted from an article by Kristin Butcher and Anne Piehl in *Journal of Policy Analysis and Management*. [↑](#footnote-ref-1)