Roughly Outlined Solutions for SØK 3001 Spring Exam 2018

Costanza Biavaschi

June 1, 2018

1 Exercise 1 (40 points)

Standard economic theory predicts that the demand for children is influenced by the cost of raising children. Holding other things constant, a decrease in the cost of raising children should lead to an increase in the demand for children. Whittington, Alm, and Peters (1990) provide evidence for this relationship, exploiting the fact that between 1913 and 1984 the value of child tax benefits in the U.S. increased substantially relative to estimates of the cost of raising children. Whittington, Alm, and Peters (1990) claim a large positive effect of child tax benefits on fertility using time series methods. Their key conclusion is based on the following equation, estimated for the period 1913 to 1984:

$$\begin{aligned} \text{Fertility Rate}_t = &\beta_0 + \beta_1 \text{Personal Exemption}_t + \beta_2 \text{Male and Asset Income}_t + \\ &+ \beta_3 \text{Unemployment}_t + \beta_4 \text{Infant Mortality}_t + \beta_5 \text{Immigration}_t + \\ &+ \beta_6 \text{Female Wage}_t + \beta_6 \text{Pill}_t + \beta_8 \text{WW2}_t + \beta_9 \text{Time Trend}_t + u_t. \end{aligned} \tag{1}$$

Here Fertility Rate_t measures the number of children born per 1,000 women; Personal Exemption_t is the dollar value of the personal tax exemption, that is, the dollar amount that a resident taxpayer is entitled to claim as a tax deduction in the presence of dependent children; Male and Asset Income_t is the dollar value of personal income per family, net of female earnings; Unemployment_t measures the share of people who are unemployed; Infant Mortality_t measures the number of children who die per 1,000 live births; Immigration_t measures the share of people who are foreign born; Female Wage_t is the dollar value of after-tax female wage; Pill_t is a dummy variable that equals one in years 1963-1984, when birth control became widely available; WW2_t is a dummy variable that equals one in years during which the US was in World War II and Time Trend_t is a time trend equal to one in 1913 and increasing by one unit each year.

In this exercise you are asked to revisit this question, discussing and interpreting the findings in Whittington, Alm, and Peters (1990) and the critique provided in Goda and Mumford (2010). All relevant results are reported in Table 1, on page 7.

(a) Interpret the regression results reported in column (1).

10 possible points. A large number of students misinterpreted the ratio variables. A number of them performed t-tests that however were not required. The following penalties were applied:

- **Ratios misinterpreted/some vagueness (no unit of measure or imprecise at places): 8 points**
- **Ratios wrong plus other misinterpretation (eg % in the wrong places, not using the unit f measurement): 5 points**
- **Vague answer, but an answer present: 2 points**

Solution. This question tests some basic ability to interpret regression results. Variables of various types are included in the model (dummies, rates, shares...). It is important that students highlight the ceteris paribus interpretation. All other things being equal, the following holds: a ten dollar increase in personal exemption is expected to increase the number of children born per 1,000 women by almost two children (1.78). A 1,000 dollars increase in family income and assets is expected to increase fertility by more than 3 children (3.5); a 1% increase in the unemployment rate is expected to decrease fertility by 0.68 (so roughly a 2% increase reduces the n. of children born by 1); roughly about every 3 children who die we expect the n. of children born to increase by slightly more than one (1.17); a one percent increase in the foreign born is expected to increase fertility by more than 9 extra children; a one extra dollar of hourly wage is expected to increase the n. of children born by 15; the availability of birth controls has reduced the n. of children born by 25 children per 1,000 women; during the War, the n. of children born was smaller by 29 births per 1,000 women; finally, we experienced throughout the century a downward trend in ferility: almost a reduction of one child for each year between 1913 and 1984. As usual the intercept has little meaning: it would capture the average fertility when all the regressors are zero.

(b) Column (1) reports results of a simplified version of column (2). Explain the restrictions that are imposed in column (1) compared to column (2). Explain the consequences on the model in column (1) if these restrictions are violated. Finally, explain how you could test these restrictions.

4 possible points. Most students did not understand this question. A large number simply listed assumptions of the OLS estimator, without realizing the one that was key here. If serial correlation was mentioned, full credit was given. The following penalties were applied:

- **Vague answers: 2 points**

Solution. This question tests students' knowledge of different estimators (knowing what OLS vs FGLS is) but especially asks to discuss issues of serial correlation. It is important that the students realize that, if other assumptions hold, the OLS will still be unbiased but inefficient. There must be a discusion of the effect on the variance of the presence of serial correlation and understanding of how to test for it. The estimation in column (2) employs a Feasible Generalized Least Squares estimator. This is motivated by a stated concern about serial correlation in the error term u_t in equation (1).

Consider the baseline model of equation (1). Now suppose that:

$$u_t = \rho u_{t-1} + e_t \quad \text{with } e_t \quad iid$$

The model in column (1) imposes $\rho = 0$. Under this condition and the additional assumption TS1-TS6 (should be explained), the OLS estimator would be unbiased and consistent.

In the presence of serial correlation (i.e. $\rho \neq 0$), the OLS is no longer BLUE, the OLS standard errors and test statistic are not valid even asymptotically. The students could show that the variance of the OLS estimator would take the form (chapter 12.1b):

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{SST_x} + 2\frac{\sigma^2}{SST_x^2}\sum_{t=1}^{n-1}\sum_{j=1}^{n-t}\rho^j x_t x_{t+j}.$$

The second term would not be accounted for. A large ρ and positive correlation over time of the regressors can cause the bias of the OLS variance estimator to be large.

It is always advisable to test for serial correlation (chapter 12.2a) - note, we have not discussed in class the Durbin Watson test, so a discussion of that test is not expected.

A test for serial correlation assumes that $E(e_t|u_{t-1}, u_{t-2}, \ldots) = 0$ and $V(e_t|u_{t-1}) = \sigma_e^2$. The hypothesis are:

$H_0 : \rho = 0$
$H_1 : \rho \neq 0$

Under the null hypothesis $\{u_t\}$ are weakly dependent so we could use a standard t-test. However, the error term is unobserved. Hence, in a first step we ned to estimate the main model and obtain residuals, then we can regress the residuals on their one-period lag. The t-statistic takes then the usual form.

Finding serial correlation motivates the use of the FGLS.

(c) Compare and contrast the estimators used in column (1) and column (2).

4 possible points. Most students did not understand this question. A large number simply compared the estimates rather than the estimators. That cannot be accepted. Partial credit was given to those that discussed heteroskedasticity, although the question SPECIFICALLY excluded this answer. The following penalties were applied:

- **Discusses FGLS in the context of heteroskedasticity: 3 points**
- **Vague answer that compares estimates: 1 point**

Solution. This test understanding of different estimators. The FGLS does not assumes serial uncorrelation and it is based estimating by ordinary least squares a modified model of the type:

(Fertility $\text{Rate}_t - \hat{\rho}$Fertility Rate_{t-1}) on (Personal $\text{Exemption}_t - \hat{\rho}$Personal Exemption_{t-1}) and so on for all regressors (one would also need to correct the estimation to include the first period). The OLS applied to this quasi-differentiated data will deliver t-stat and F-stat that are asymptotically valid. Note that FGLS will only be asymptotically valid and in small samples will be biased. Big differences between OLS and FGLS should be taken with suspicion. Full derivations are not expected but these issues have been discussed in class and have also been discussed in the Discussion Board on Blackboard.

(d) For Whittington et al. (1990) also estimate a model in which they include a new variable, which they call *Triangular Personal Exemption*, constructed in the following way:

$$\text{Triangular Personal Exemption}_t = \frac{1}{3}\Big(\text{Personal Exemption}_t + \text{Personal Exemption}_{t-1} + \\ + \text{Personal Exemption}_{t-2}\Big)$$

Does this variable capture the long run impact of personal exemptions on fertility? Why could it be of interest to include Triangular Personal Exemption_t in the model?

2 possible points. Most students did not understand this question. Little weight was given to it. As long as there was some discussion of how lags can measure long term effects full credit was given. The following penalties were applied:

- **No answer: 0 points**
- **Some discussion of how lags measure long term propensity: 2 points**

Solution. This question tests the understanding of LRP and, somewhat more challenging, hints to issues of seasonality. Model (1) and (2) do not take into account that, due to biological constraints, the birth of a child will lag the decision to have a child. Rather, model (1) and (2) assume an immediate effect of tax credits on fertility. It seems natural to include lags in the baseline model:

$$\begin{aligned}\text{Fertility Rate}_t =& \beta_0 + \beta_1 \text{Personal Exemption}_t + \beta_2 \text{Personal Exemption}_{t-1} + \\ &+ \beta_3 \text{Personal Exemption}_{t-2} + \ldots + u_t\end{aligned}$$

In such model the long-run propensity can be found by assuming a steady state so that:

$\text{Personal Exemption}_t = \text{Personal Exemption}_{t-1} = \overline{\text{Personal Exemption}}$.

Hence, $LRP = \beta_1 + \beta_2 + \beta_3$. In the estimated model we rather have that:

$$\text{Fertility Rate}_t = \beta_0 + \theta \text{Triangular Personal Exemption}_t + \ldots + u_t$$

The two things are not equivalent. With the triangular structure we are averaging three consecutive values of the time series and, in order to avoid a shift in the level, divided by 3. This eliminates "seasonal variation", in this case primarily stemming from biological constraints.

(e) Zhang et al. (1994) mention that there is a concern that some series in the Whittington et al. (1990)'s study may be non-stationary. However, Zhang et al. dismiss these criticisms claiming that because "the time trend is insignificant" in their estimation, there is no concern that the results are being driven by a regression of "time against time". Using results in column (3), show that the time trend is insignificant and then discuss whether you agree or disagree with this statement. Discuss this issue also comparing results with those presented in column (4).

8 possible points. Most students did not discuss stationarity and all tested the significance of the time trend. The following penalties were applied:

- **Only trend tested: 4 points**
- **Trend tested and some attempt to discuss properties of series/models (e.g. understanding of first-difference models, but no mention of unit root): 6 point**

Solution. This question tests a core concept: unit roots. Students should highlight the difference between trends and unit roots. We can test whether the time trend is insignificant by checking:

$H_0 : \beta_9 = 0$
$H_1 : \beta_9 \neq 0$

A standard t-test requires:

$$t_{stat} = \frac{\hat{\beta}_9 - 0}{se(\hat{\beta}_9)} = \frac{-0.377}{4.71} = 0.08$$

Under H_0 the t-statistic is distributed as a T with $n - k - 1 = 68 - 9 - 1 = 58$ degrees of freedom. The critical values are (roughly) 1.67 at 10%, 2 at 5% and 2.66 at 1% significance levels. The t-statistic is below any of these values so we fail to reject the null hypothesis and conclude that we do not have enough evidence that suggest the presence of a time trend in fertility rates.

However, an insignificant time trend coefficient estimate does not alleviate concerns that some of the series are I(1). It is extremely important not to confuse trending and highly persistent behaviors. A series can be trending but not highly persistent and can be highly persistent and trending. In addition, if the series are I(1) testing on a time trend would be invalid. In fact, suppose that:

$$\text{Fertility Rate}_t = \beta_0 + \beta_1 \text{Personal Exemption}_t + \ldots + \text{Fertility Rate}_{t-1} + u_t$$

For the sake of the argument, let's disregard all the observable characteristics.

$$E(\text{Fertility Rate}_t) = E(\text{Fertility Rate}_{1913})$$
$$V(\text{Fertility Rate}_t) = t\sigma_u^2$$

The variance of a random walk increases as a linear function of time. This process cannot be stationary and violates the key assumptions of the CLM.

Granger and Newbold (1974) in their seminal paper on spurious regressions argue that annual macro series, like those used in this study, are almost always I(1); thus, regressions involving the levels will be misleading, suggesting relationships when there may be none.

A unit root test, rather than a test on the trend, seems necessary. This was carried out in class. There is not enough information in the table to carry it out, but the fact that the estimates in first differences show a statistically insignificant effect on fertility (a big change compared to before!) hints to the first few models being poorly specified.

(f) Goda and Mumford (2010) examine additional features of the tax code that provide tax subsidies to families with children. Rather than focusing only on the personal tax exemption, they construct a new variable, *total child tax subsidy*, that is the sum of personal exemption and additional child tax benefits, namely child tax benefits from the earned income tax credit, the child and dependent care tax credit, and the child tax credit Their results are reported in column (5) of Table 1. Discuss possible reasons that explain why results differ between column (4) and (5).

8 possible points. Only one student related to measurement error. Most students brought up issues of correlation across variables (multicollinearity at times). This was accepted as an answer, if well argued. The following penalties were applied:

- **Discusses correlation across variables: 7 points**
- **Vague answer: 3 points**

Solution. This question is about measurement error. We have not discussed it in the time series context, but students could list the strong TS.1-TS.6 assumptions and reproduce what learned for cross-sectional data. Good understanding would further discuss that attenuaton bias will occur in an otherwise well-specified model. The dynamics in column (5) are still not well-specified The models in column (1)-(4) might suffer of measurement error in Personal Exemption. While measurement error in the dependent variable does not have many consequences, that of the independent variable results in biased estimates of the effects of interest. The students should show this (done in class, in the textbook and in the tutorial).

(g) Conclude this exercise by summarizing the results in Table 1 and drawing conclusions on the relationship between tax incentives and fertility. In writing your conclusions, pay particular attention to discussing the statistical and economic significance of the relationship between child benefits and tax benefits across models. What is your preferred model among those presented? Could one further improve on your preferred model?

4 possible points. Most students picked model 3 as their favorite, missing the issue of unit roots. Some partial credit given to vague answers. The following penalties were applied:

- **Picks model 3: 2 points**
- **Vague answer: 1 point**

Solution. This question makes sure that students have understood the logic of the question and of the steps of an empirical analysis. In the exercise we have first shown that the main result from column (1), namely the strong results reported in Whittington et al. (1990), are not robust to first-differencing. Column (1) showed that $100 in tax benefits (in 1967 dollars) are associated with an increase in the general fertility rate of 11.6. Moving from Column (3) to Column (4) illustrates the effect of first-differencing the series that contain a unit root. The coefficient on the tax subsidy flips sign and decreases in magnitude. Column (5) repeats the analysis including other child tax benefits in the tax subsidy series. While the coefficient on the total child tax subsidy variable are of the same signs as in Columns (4), it is no longer significant and smaller in magnitude. Because of the increasing importance of tax subsidies for children other than the personal exemption and their more salient nature, the changes in the key coefficients that result from adding in these other tax benefits cast additional doubt regarding the true effect of tax subsidies on fertility. Overall, Table 1 shows that Whittington et al.'s result is sensitive to correcting for unit roots by first-differencing and adding the tax subsidies for children in other parts of the tax code. The model in column (5) still disregards any dynamic effect of tax exemption on fertility. This is a limitation of this analysis and one could include further lags of the key explanatory (or dependent!) variable and eventually test whether such model outpeforms the static one.

Table 1: Child Tax Benefits and Fertility

	OLS	FGLS	FGLS	First Difference	First Difference
	(1)	(2)	(3)	(4)	(5)
Personal Exemption	0.178	0.121	-	-0.084	-
	(0.0977)	(0.0446)	-	(0.042)	-
Triangula Personal	-	-	0.191	-	-
Exemption	-	-	(0.0477)	-	-
Total Child Tax Subsidy	-	-	-	-	-0.007
	-	-	-	-	(0.006)
Male and Asset Income	0.0035	-0.0004	-0.0004	-0.003	-0.001
	(0.0031)	(0.0027)	(0.0027)	(0.002)	(0.000)
Unemployment	-68.12	-73.43	-36.800	-20.985	-8.957
	(25.818)	(34.20)	(39.60)	(31.280)	(27.301)
Infant Mortality	0.393	0.083	0.303	-0.042	-0.054
	(0.321)	(0.255)	(0.3817)	(0.315)	(0.274)
Immigration	964.13	774.24	1529.2	68.878	194.315
	(329.44)	(311.31)	(480.44)	(119.073)	(138.742)
Female Wage	15.427	5.647	-2.157	7.472	1.924
	(5.286)	(15.686)	(14.098)	(5.792)	(1.196)
Pill	-25.383	-10.856	-8.958	-1.91	-0.44
	(11.961)	(6.126)	(5.522)	(1.020)	(0.841)
WW II	-29.419	-17.223	-5.353	5.138	3.468
	(8.057)	(4.989)	(3.947)	(3.377)	(2.572)
Time Trend	-0.843	-0.539	-0.377		
	(0.543)	(0.538)	(4.71)		
Intercept	55.944	102.979	81.628	-0.618	-1.174
	(25.831)	(24.666)	(32.251)	(0.954)	(0.943)
N	72	71	68	71	71
R^2	0.829	0.916	0.941	0.203	0.103

In all columns, the dependent variable is the fertility rate at time t, which measures the number of children born per 1,000 women *Personal Exemption* is the dollar value of the personal tax exemption, that is the dollar amount that a resident taxpayer is entitled to claim as a tax deduction if one has dependent children; *Triangular Personal Exemption* is defined in the text; *Male and Asset Income* is the dollar value of personal income per family net of female earnings; *Unemployment* measures the share of people who are unemployed; *Infant Mortality* measures the number of children who die per 1,000 live births; *Immigration* measures the share of people who are foreign born; *Female Wage* is the dollar value of after tax female wage; *Pill* is a dummy variable that equals one in years 1963-1984; *WW2* is a dummy variable that equals one in years during which the US was in World War II and *Time Trend* is a time trend equal to one in 1913 and increasing by one unit each year.

2 Exercise 2 (40 points)

Classical criminology assumes that criminals are rational beings who weigh the costs and benefits of their actions. Gary Becker (1968) produced the first fully fledged theory of crime based on rational behavior. His research led to an upsurge of interest in the economics of criminal behavior. One of the central predictions of Becker's theory is that crime will decrease when police presence increases. A basic problem with this prediction is that it largely failed to find empirical support during the 1970s and 1980s. In a survey of the literature, Samuel Cameron (1988) reports that in 18 out of 22 papers surveyed researchers found either a positive effect of police presence on crime or no relationship between these variables.

Most of these studies estimated models in which the number of violent crimes in city c at time t were regressed on the number of policemen per capita in city c at time t.

In this exercise you are asked to revisit this question, discussing and interpreting the findings in Levitt (1997). The sample includes 122 cities observed between 1975 and 1995. Levitt revisits previous evidence and estimates several versions of the following model:

$$\begin{aligned}\log \text{Violent Crime}_{ct} =& \beta_0 + \beta_1 \log \text{Police per Capita}_{ct} + \beta_2 \log(\text{State prisoners per capita})_{ct} + \\ &+ \beta_3 \text{Unemployment rate}_{ct} + \beta_4 \text{State income per capita}_{ct} + \\ &+ \beta_5 \text{Effective abortion rate}_{ct} + \beta_6 \log(\text{City population})_{ct} + \\ &+ \beta_7 \text{Percentage black}_{ct} + u_{ct}. \end{aligned} \quad (2)$$

Here $\log \text{Violent Crime}_{ct}$ is the log of per capita city crime at time t, $\log \text{Police per Capita}_{ct}$ and $\log(\text{State prisoners per cap}$ are the log of police per capita and of the number of prisoners in the State in city c at time t, respectively; $\text{Unemployment rate}_{ct}$ measures the shares of unemployed in the population in city c at time t; $\text{State income per capita}_{ct}$ measures State income per capita in 10,000 dollars; the effective abortion rate (in 100) is the weighted average of the abortion rate of crime-aged individuals; $\log(\text{City population})_{ct}$ is the log of the population in city c at time t and $\text{Percentage black}_{ct}$ is the percentage of African American individuas in city c at time t. All relevant results are reported in Table 2, on page 11.

(a) Consider first the model in column (1) of Table 2. Discuss under which conditions this model might pin down the causal impact of police on crime and discuss whether these conditions are likely to hold in this context.

20 possible points. Most students presented a broad discussion of OLS assumptions, sometimes in the time series context rather than in the pooled cross-section one. Very few students were able to relate the identification conditions to the exact problem under study. The following penalties were applied:

- **Broad discussion: 5 points**
- **Broad discussion + 5 points each for mentioning simultaneity, OVB, measurement error**

Solution. This question tests the understanding of OLS assumptions and whether students can critically evaluate an empirical problem. Column (1) employs the OLS estimator on the following model:

$$\begin{aligned}\log \text{Violent Crime}_{ct} =& \beta_0 + \beta_1 \log \text{Police per Capita}_{ct} + \beta_2 \log(\text{State prisoners per capita})_{ct} + \\ &+ \beta_3 \text{Unemployment rate}_{ct} + \beta_4 \text{State income per capita}_{ct} + \\ &+ \beta_5 \text{Effective abortion rate}_{ct} + \beta_6 \log(\text{City population})_{ct} + \\ &+ \beta_7 \text{Percentage black}_{ct} + u_{ct}\end{aligned}$$

Indicate with $\boldsymbol{X}_{ct}$ the set of regressors in the previous equation. The key assumption for causal identification of the impact of police on crime is $E(u_{ct}|\boldsymbol{X}_{ct}) = 0$. A violation of this condition would result in a biased $\hat{\beta}_1$. In fact, the second term of this summation would not be zero:

$$E(\hat{\beta}_1) = \beta_1 + \frac{Cov(\log \text{Police per Capita}_{ct}, u)}{Var(\log \text{Police per Capita}_{ct})}$$

The fact that the OLS estimator delivers a positive relationship rather than a negative one, seems to suggest that:

$$\frac{Cov(\log \text{Police per Capita}_{ct}, u)}{Var(\log \text{Police per Capita}_{ct})} > 0$$

Possible reasons for this to happen are:

- Simultaneity bias: if more police is hired when crime increases, a positive correlation between crime and police can emerge even if police reduces crime.
- The presence of unobserved heterogeneity across cities will impart an upward bias on cross-sectional estimates of police effectiveness. Cities that have a high level of underlying criminality are likely to have both high crime rates and large police forces. Detroit has twice as many police per capita than Omaha, and violent crime over four times as high, but it would be a mistake to attribute the differences in crime rates to the presence of the police.
- A final source of bias is that as police increases reporting rates might also increase if police workload (i.e. more police) is reduced.
- Finally, measurement error per se should not be a problem as it would bias towards zero the results but not make the estimated effects positive.

(b) Test whether permanent differences in crime rates across cities are important in explaining violent crimes. Explain whether a model that includes such differences is sufficient to identify the causal effect of police on crime.

10 possible points. Some students did not perform the test. Full credit was given even if the degrees of freedom were incorrect. The following penalties were applied:

- **No test and some vague discussion: 5 points**

Solution. This question requires students to understand the role of fixed effects and tests basic knowledge of an F-test. The model in column (2) applies the OLS to the following equation:

$$\begin{aligned}\log \text{Violent Crime}_{ct} = &\beta_0 + \beta_1 \log \text{Police per Capita}_{ct} + \beta_2 \log(\text{State prisoners per capita})_{ct} + \\ &+ \beta_3 \text{Unemployment rate}_{ct} + \beta_4 \text{State income per capita}_{ct} + \\ &+ \beta_5 \text{Effective abortion rate}_{ct} + \beta_6 \log(\text{City population})_{ct} + \\ &+ \beta_7 \text{Percentage black}_{ct} + \delta_1 city1 + \delta_2 city2 + \ldots + \delta_{121} city121 + u_{ct}\end{aligned}$$

Here $city1 - city121$ are indicators that take value of one in city c and zero otherwise. These variables are able to capture permanent difference in crime rates across cities. This model therefore tackle the second challenge to exogeneity listed in the previous section.

To test whether permanent differences in crime rates across cities are important in explaining violent crimes, we perform a joint significance test on these variables.

$H_0: \delta_1 = \delta_2 = \ldots = \delta_{121} = 0$

$H_1:$ not H_0

The test statistic is:

$$F - stat = \frac{R_u^2 - R_r^2}{(1 - R_u^2)} \times \frac{n-k-1}{q} = \frac{0.93 - 0.601}{1 - 0.93} \times \frac{2005 - (121+7) - 1}{121} = 72.86$$

Under the null hypothesis, the F-statistic is distributed as an $F_{121;1876}$. The critical values are: at 10% significance level the critical value is 2.30, at 5% is 3 and at 1% is 4.61. The F-statistic is well above all these critical values so we reject the null hypothesis and city heterogeneity enters the model.

Such model would solve for the presence of unobserved heterogeneity and capture all time-invariant omitted variables. Nonetheless it would still not account for the potential simultaneity bias, or capture all those time-varying factor that are correlated with both police and crime and not captured by the other variables.

(c) Define the problem of weak instruments and discuss whether you think that the model exhibits a weak instrument problem.

10 possible points. Most students had a discussion of weak instruments, not all performed the test. Most mentioned the rule of thumb. The following penalties were applied:

- **No test and some vague discussion: 5 points**
- **Discussion + test (on the correct coefficient): 8 points**
- **Discussion + test (on the correct coefficient+ rule of thumb): 10 points**

Solution. This question requires students to discuss potential problems with IV strategies. We saw in class that

$$plim \hat{\beta}_1^{IV} = \beta_1 + \frac{Corr(Firefighters, u)}{Corr(Firefighters, Police)} \frac{\sigma_u}{\sigma_{police}}$$

This equation highlights that if $Corr(Firefighters, Police)$ is small, even in case of a close to zero $Corr(Firefighters, u)$, we could have a large bias in the IV. In this case, comparing column (2) with column (3) we find that the IV is 6 times bigger than the OLS. A 10% increase in police per capita is expected to decrease violent crimes by almost 5% (other things being equal). The OLS estimator suggested that a 10% increase in police would have lowered violent crimes by a mere 0.7%. These seem substantial differences. However, the instrument will pass the rule of thumb of being greater than 3.2. The test is performed on the first-stage regression:

$$\begin{aligned}\log \text{Police per Capita}_{ct} = & \pi_0 + \pi_1 \log \text{Firefighters per Capita}_{ct} + \\ & + \pi_2 \log(\text{State prisoners per capita})_{ct} + + \pi_3 \text{Unemployment rate}_{ct} + \\ & + \pi_4 \text{State income per capita}_{ct} + \\ & + \pi_5 \text{Effective abortion rate}_{ct} + \pi_6 \log(\text{City population})_{ct} + \\ & + \pi_7 \text{Percentage black}_{ct} + \delta_1 city1 + \delta_2 city2 + \ldots + \delta_{121} city121 + u_{ct}\end{aligned}$$

Here we test:

$H_0: \pi_1 = 0$

$H_1: \pi_1 \neq 0$

This is a standard t-test which requires:

$$t_{stat} = \frac{\hat{\pi}_1 - 0}{se(\hat{\pi}_1)} = \frac{0.206}{0.050} = 4.12$$

Under H_0 the t-statistic is distributed as a T with $n - k - 1 = 2005 - 121 - 1 = 1883$ degrees of freedom. The critical values are (roughly) 1.64 at 10%, 1.96 at 5% and 2.57 at 1% significance levels. The t-statistic is above all critical values and above the rule of thumb, so we reject the null hypothesis and conclude that the presence of firefighters is a good predictor for the number of policemen in a city. Of course, it could still be that our instrument is relevant but invalid, i.e. violates the identifying assumption $Corr(Firefighters, u) = 0$. This is untestable with only one instrument, but a student could think of reasons why this condition might be violated in the model above.

Table 2: The Impact of Police on Crime

	OLS	OLS	OLS	IV
	log(Violent Crime)	log(Violent Crime)	log(Police per capita)	log(Violent Crime)
	(1)	(2)	(4)	(3)
log(Firefighters per capita)	-	-	0.206	-
	-	-	(0.050)	-
log(Police per capita)	0.562	-0.076	-	-0.435
	(0.056)	(0.061)	-	(0.231)
log(State prisoners per capita)	0.25	-0.131	-0.077	-0.171
	(0.039)	(0.036)	(0.022)	(0.044)
Unemployment rate	3.573	-0.741	0.265	-0.48
	(0.473)	(0.365)	(0.314)	(0.404)
State income per capita	0.05	-0.003	0.211	0.003
(x10,000)	(0.005)	(0.006)	(0.005)	(0.007)
Effective abortion rate	-0.214	-0.15	0.045	-0.141
(x100)	(0.045)	(0.023)	(0.026)	(0.025)
log(City population)	0.072	0.203	-0.014	0.178
	(0.012)	(0.063)	(0.047)	(0.067)
Percentage black	0.627	0.233	0.493	0.398
	(0.074)	(0.334)	(0.264)	(0.345)
City dummies and year dummies?	Only year	yes	yes	yes
R^2	0.601	0.93	0.962	-
N	2005	2005	2005	2005

The dependent variables are reported in the first row on top of each column, with $\log \text{Violent Crime}_{ct}$ being the log of per capita city crime at time t and $\log \text{Police per Capita}_{ct}$ being the log of police per capita in city c at time t. log(Firefighters per capita) is the log of the number of firefighters per capita; log(State prisoners per capita)$_{ct}$ is the number of prisoners in the State in city c at time t, respectively; Unemployment rate$_{ct}$ measures the shares of unemployed in the population in city c at time t; State income per capita$_{ct}$ measures State income per capita in 10,000 dollars; the effective abortion rate (in 100) is the weighted average of the abortion rate of crime-aged individuals; log(City population)$_{ct}$ is the log of the population in city c at time t and Percentage black$_{ct}$ is the percentage of African American individuas in city c at time t.

3 Exercise 3 (20 points)

Throughout the course we have analyzed linear models of the form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u.$$

Albeit linear relationships are very powerful, not all economic models are linear.

Hence, consider now a non-linear specification in which y is a generic, known, non-linear function $f(\cdot)$ of characteristics and parameters:

$$y = f(x_1, \ldots, x_k; \beta_0, \beta_1, \ldots, \beta_k) + u.$$

Data are assumed to be i.i.d, $E(u|x_1, \ldots, x_k) = 0$ and $V(u|x_1, \ldots, x_k) = \sigma^2$.

(a) Define a method that you could use to estimate the vector of parameters $\beta_1, \ldots, \beta_k$. Note: you do not have do derive these parameters, only to describe a method for doing it.

10 possible points. Some students thought of using MLE and referred to Probit/Logit models. This was accepted. Other students The following penalties were applied:

- **Minimizing SSR/MLE: 10 points**
- **Vague discussion on some estimator: 5 points**

Solution. This question requires students to understand the basics of the least squares estimation and apply it to a new context. Analogous to the OLS which minimizes:

$$\min_{\beta_1,\ldots,\beta_k} Q(\beta_1, \ldots, \beta_k) = \min_{\beta_1,\ldots,\beta_k} \sum_{i=1}^{n} \Big(y_i - \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k\Big)^2$$

One could think of estimating the parameters using an analogous least-squares estimation:

$$\min_{\beta_1,\ldots,\beta_k} Q(\beta_1, \ldots, \beta_k) = \min_{\beta_1,\ldots,\beta_k} \sum_{i=1}^{n} \Big(y_i - f(x_1, \ldots, x_k; \beta_0, \beta_1, \ldots, \beta_k)\Big)^2$$

The first order condition with respect to β_1 for example would take the form:

$$\frac{\partial Q(\beta_1, \ldots, \beta_k)}{\partial \beta_1} = \sum_{i=1}^{n} \frac{\partial f(x_1, \ldots, x_k; \beta_0, \beta_1, \ldots, \beta_k)}{\partial \beta_1} \Big(y_i - f(x_1, \ldots, x_k; \beta_0, \beta_1, \ldots, \beta_k)\Big) = 0$$

It is easy to see that the linear model is a special case of this more generic situation, as the relative first order condition was:

$$\frac{\partial Q(\beta_1, \ldots, \beta_k)}{\partial \beta_1} = \sum_{i=1}^{n} x_i \Big(y_i - f(x_1, \ldots, x_k; \beta_0, \beta_1, \ldots, \beta_k)\Big) = 0$$

(b) Let $\hat{\beta}_0, \ldots, \hat{\beta}_k$ indicate the estimator resulting from part (a). Describe which statistical properties you would like this estimator to exhibit. [Note: again, you are not required to make derivations]

5 possible points. Most students confused assumptions with statistical properties. The following penalties were applied:

- **Vague answers: 2**

Solutions. This question is a test of basic knowledge. As we have studied throughout the course, unbiasedness and large sample consistency are two properties that econometricians consider desirable. Unbiasedness requires:

$$E(\hat{\beta}_j) = \beta_j \quad j = 1, \ldots, k$$

This properties requires the estimator to be centered around the true parameter. In other words, while each single realization might differ from the true β_j in the population, on average an unbiased $\hat{\beta}_j$ would deliver the true population parameter.

Consistency requires:

$$plim \quad \hat{\beta}_j \rightarrow \beta_j$$

That is, in large sample, the probability limit of $\hat{\beta}_j$ is β_j.

(c) Explain how you would check *empirically* that the estimator you have proposed at point (a) satisfies the properties that you have discussed at point (b).

5 possible points. Most students discussed testing of the assumptions with statistical properties. The following penalties were applied:

- **Vague answers: 2**

Solution. This question links to the Monte Carlo studies that we have used in class to study the properties of the OLS estimator. It is a "reward" to those students who have fully studied the course material and understood the wide application of the concepts covered. One way to show the behavior of an estimator is to perform Monte Carlo simulations. We have used this tool several times during the course, e.g. to discuss unbiasedness of the OLS under the CLM assumptions and when those assumptions were violated. Similarly here, one could use Monte Carlo simulations to study the behavior in small and large samples of $\hat{\beta}_j$.

Monte Carlo simulation is a method of analysis based on artificially recreating a random process with a computer, performing the estimation, observing the results and running the whole procedure many times.

One could generate the data for x_1, $x_2, \ldots, x_k$ and u as normal random variables. One could then set values for the parameters $\beta_0, \ldots, \beta_k$ and derive then y. One could then apply the estimator proposed at (a), collecting the results. By repeating this procedure r times and looking at the average of the results one could check unbiasedness.

Consistency could be studied by looking at the large sample behavior of the estimator, i.e. increasing the sample size. One could also use a data generating process which violates some of the assumptions made and study the asymptotic behavior of the estimator.