Instructions: Write succinctly. Each problem presents a task and additional questions.

1. The LIBOR stands for the London interbank offered rate, which is a daily average of the interest rate that banks charge other banks to lend money. For example, in the market for interbank loans denominated in US dollars, 18 banks contribute an interest rate every morning to the British Bankers Association (BBA). The BBA throws out the highest 4 and the lowest 4 interest rates, and averages the middle 10, creating a trimmed mean of the interest rates. This mean is the LIBOR.

In 2008 in the height of the financial crisis it became apparent than the LIBOR did not reflect the true state of uncertainty in the economy and suspicions for fraud started to arise. The LIBOR seemed to be low, giving the impression that banks could easily borrow funds from one another and that banks were healthy. In hindsight, it is well known that many banks had liquidity problems and could not have been as confident as they appeared to be. Banks likely colluded in a cartel, in order to drive the trimmed mean up or down.

You are an analyst conducting an investigation into the manipulation of the LIBOR. You have obtained sensitive information on daily interest rate offers by banks. You consider several approaches to detect LIBOR manipulation. You decide to use tests with 1 percent significance level (unless otherwise specified).

a. You want to determine whether banks in 2008 were colluding by submitting similar low interest rates. If there is a cartel, then the offers should be nearly identical, generating a low variance. You compare a given day in 2008 to a day in 2006, see the variables **Offer 2008** and **Offer 2006** in Table 1. To determine whether offered interest rates were too similar, you want to check the variances of the offered rates with an F-test. Interpret your results. Why does the similarity in offers indicate colluding?

b. Following your intuition from above, you are now interested in determining which banks were involved in the cartel. You collect information on the offers of the 18 banks and you look at a correlation matrix in Table 2, which shows the correlation between any two banks' interest rate offers. Banks have been anonymized by replacing their true names with letters. You assume that colluding banks will issue identical interest offers. Which banks are likely involved? How did you determine which banks are involved?

c. To determine whether banks are unusually confident during a time of crisis, you want to see whether the trimmed mean of the offers is smaller in 2008 than in 2006. If banks did indeed feel the crisis, then the interest rates in 2008 should be higher. Conduct a t-test of the variables **Libor 2006** and **Libor 2008** from Table 1, assuming unequal variances. Interpret your results.

d. You realize that having only 10 observations per sample, the t-test from point c might not be appropriate, so you decide to conduct a Sign test on the data in Table 3. What is the outcome of the test and what do you learn about the distribution of the offers? Does the test give the same result as the test in the previous question?

e. You are going deeper into the investigation and you want to see how much the involved banks profited. You collect data on daily profits for **involved banks** and **uninvolved banks** and you run a 2-sided t-test with unequal variance, picking a 1 percent significance level. However, the output of the test is not complete in Table 4.

The variables' summary statistics are in Table 1. Interpret the result in Table 4. Then find the missing confidence interval of the difference in means.

f. You decide to employ a chi-squared test and check the application of Benford's Law. According to this law, in many naturally occurring numerical data sets the first and second digits of any given number follow a specific distribution. You take the second digit of Libor from June 2007 to November 2007 and you want to compare their distribution to the expected distribution according to Benford's Law. See Table 5. Determine whether the frequency of second digits observed in the data is significantly different than the expected frequency.

g. You are wondering whether the risk exposure of banks determines their participation in the cartel for different years. You perform an ANOVA test, where the response variable is an indicator variable for cartel participation. The populations differ by the two factors: **risk** exposure ratio and **year** of observation. You get the output from Table 6. How many years are in the sample? Interpret the output considering a test with 10 percent significance.

2. A central idea in conflict research is that income and the onset of civil wars are related. People in countries with low income experience poverty, which can lead to the onset of armed conflict in order to redistribute scarce resources. Conflict destroys both infrastructure and life, leading to smaller possibilities of generating income and, therefore, to lower income. Lower income then leads to more conflict, and so the vicious circle of poverty and war closes.
Researchers have been trying to estimate causal effects for the impact of income on conflict, in order to establish when a country needs foreign aid and how best to keep societies secure. The vicious circle of poverty and war spells out the main problem of reverse causality, which impedes researchers from obtaining credible causal estimates.
You, as an analyst in the World Bank, have decided to take a different approach. You want to develop a method to forecast the effect of income on conflict onset and for that purpose you have obtained the dataset of Burke (2012), which gives you many useful variables to study your problem. The dataset covers 150 countries and it spans from 1961 to 2000. You observe the following variables:

Conflict – Indicator variable denoting the occurrence of conflict in a given year. It takes a value 1 for conflict events, and 0 otherwise. Interpreted as the probability of conflict.
GDP capita – Gross Domestic Product per capita in a given year and country divided by 1000 USD. This is a measure of income.
Share of young – Share of the population under 15 years old
Share of old – Share of the population aged 65 years and above
Share of urban – Share of urban population
Exports – Exports as a percentage of GDP
Temperature – Average temperature

You try out different approaches and present your analysis in Table 9. Summary statistics are presented in Table 7 and Table 8.

a. In column 1 of Table 9 you have estimated a simple regression. Write down the estimated regression equation Write down two point predictions. In the first point prediction use the mean values for the variables in the model. In the second point prediction increase the mean value of **GDP capita** by one percent. Take the difference

of the two point predictions and derive the effect of a one percent increase in the mean of **GDP capita** on the probability of **conflict**.

b. In column 2 of Table 9 you try a different modelling approach, by including the lag of **GDP capita** in the regression. Assume that we observe a *transitory* shock: the GDP per capita increases by 10 % this year (relative to its mean value), and it returns to its mean value for the next year. What would be the effect on the probability of **conflict** this year (t) and on the probability of **conflict** next year (t+1) if we observe the transitory shock this year (t)? Create one point prediction for the probability of conflict in the absence of shocks and two point predictions for the effect of the shock. Derive the effects. Comment on the results.

c. In column 3 of Table 9 you estimate a logit model. Interpret the effect of a 10 percent increase (relative to the mean value) in **GDP capita** in year t on the probability of **conflict** in year t. Write down one point prediction for the mean and one point prediction for the effect. Derive from these point predictions the effect of 10 percent increase of **GDP capita** on **conflict**? Comment on the differences in the predictions from this point and the previous point b.

d. In column 4 of Table 9 you include additional control variables. Observing Table 8, do you think that there is a problem of multicollinearity? Why? Which variables, if yes?

e. Omitted Variable Bias. Assume that the *true* relationship between conflict and income looks like this:

$$Conflict_{it} = \gamma + \beta \times GDPCapita_{it} + \alpha \times EthnicDiversity_i + \epsilon_{it}$$

Where i is a subscript for country and t is a subscript for year. However, in column (1) of Table 9 you are trying to estimate this equation:

$$Conflict_{it} = \delta + \beta \times GDPCapita_{it} + u_{it}$$

Given that we omit the important variable EthnicDiversity, by estimating the second regression equation with OLS we introduce a bias in the estimation of the population coefficient β. The simplified formula for this bias is given by:

$$\hat{\beta} = \beta + \alpha \times \frac{Cov(EthnicDiversity_i, GDPCapita_{it})}{Var(GDPCapita_{it})}$$

Derive this formula using the fact that one of the main assumptions for obtaining an unbiased estimate of β is that $Cov(GDPCapita_{it}, \epsilon_{it}) = 0$.

f. Assume that $\hat{\beta} < \beta$. We expect that $\alpha > 0$because in societies with larger ethnic diversity it is more likely that a majority ethnicity can accumulate power to the detriment of a minority, sparking ethnical tensions and grievances. Use the formula for the omitted variable bias from point e. What can you say about the sign of the correlation between **Ethnic Diversity** and **GDP capita**? What does the inequality $\hat{\beta} < \beta$ imply about the true value of the coefficient in column 1? (The assumptions from problem e. still hold)

g. In column 6 you reason that the omitted variable bias in the effect of income will be minimized if you include country fixed effects. Why should this reasoning work? Assume that all the difference in the coefficient on **GDP Capita** between column (1) and column (5) is due to **Ethnic Diversity**. Use the formula for the omitted variable bias in point e. Find an approximation for the sign and the value of the covariance between **Ethnic Diversity** and **GDP Capita**. Use information from all available tables. (The assumptions from problem e. still hold and $\alpha > 0$)

h. Comment on the main coefficients of interest, the **GDP capita** and the lag of **GDP capita**, in Table 9 by writing down a *short* note on the effect of income on the probability of conflict, targeted to a policy maker. Max 100 words.

Table 1. Summary Statistics			
Variables	Mean	Standard Deviation	Observations
Offers 2006	3.08	0.85	18
Offers 2008	1.99	0.12	18
Libor 2006	3.09	0.56	10
Libor 2008	2.00	0.06	10
Profits involved banks (in thousand $)	20 000	500	630
Profits uninvolved banks (In thousand $)	19 000	1000	990

Table 2. Correlation matrix

Banks	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
A	1																	
B	0.6	1																
C	0.4	0.7	1															
D	0.4	0.9	0.8	1														
E	0.3	0.9	0.7	1	1													
F	0.8	0.5	0.4	0.7	0.5	1												
G	0.6	0.3	0.5	0.5	0.3	0.6	1											
H	0.5	0.1	0.7	0.2	0.7	0.9	0.3	1										
I	0.7	0.2	0	0.2	0.3	0.3	0.6	0.2	1									
J	0.8	0.5	0.5	0.7	0.5	0.5	0.2	0.7	0.3	1								
K	0.3	0.9	0.7	1	1	0.7	0.2	0.3	0.8	0.5	1							
L	0.6	0.9	0.5	0.9	0.8	0.1	0	0.8	0.9	0.2	0.8	1						
M	0.4	0.8	0.3	0.3	0.3	0.3	0.4	0.8	0.8	0.1	0.1	0.4	1					
N	0.8	0.2	0.9	0.6	0.6	0.8	0.1	0.3	0.1	0.8	0.3	0.3	0.7	1				
O	0.3	0.3	0.7	0.5	0	0	0.5	0.3	0.7	0.2	0.6	0.5	0.7	0.7	1			
P	0.9	0.4	0.7	0.4	0.1	0.3	0.5	0	0.7	0.5	0.3	0	0.2	0.1	0.2	1		
Q	0.1	0.3	0.5	0.4	0.3	0.6	0.2	0.2	0.5	0.2	0.5	0.2	0.4	0.3	0.6	0.3	1	
R	0.8	0.5	0	0.6	0.3	0.7	0.6	0.6	0.8	0.1	0.7	0.1	0.5	0.1	0.8	0.3	0.8	1

LIBOR 2006	LIBOR 2008
2.29	1.91
2.52	1.93
2.60	1.96
2.71	1.98
2.80	2.01
3.34	2.02
3.58	2.03
3.67	2.04
3.67	2.07
3.77	2.09

Table 4. R Output

```
        Welch Two Sample t-test

data:  involved and uninvolved
t = 27.706, df = 1540.6, p-value < 2.2e-16
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 [redacted]
sample estimates:
mean of x mean of y
 20045.02  18996.87
```

Table 5. Benford's Law Chi-squared Test

Second Digit	Expected frequency (Benford's Law)	Empirical Frequency
0	0.120	0.038
1	0.114	0.122
2	0.109	0.015
3	0.104	0.374
4	0.100	0.008
5	0.097	0.107
6	0.093	0.115
7	0.090	0.107
8	0.088	0.099
9	0.085	0.015
Obs.	100	100

Table 6. R ANOVA output

```
            Df Sum Sq Mean Sq F value Pr(>F)
risk         9  3.800  0.4222   2.087 0.0647
year         2  0.053  0.0263   0.130 0.8788
risk:year   13  2.874  0.2211   1.093 0.4025
Residuals   29  5.867  0.2023
```

Table 7. Summary Statistics

	Mean	SD
Conflict	0.186	0.389
GDP Capita	7.453	1.572
Share of young	35.757	9.961
Share of old	6.030	4.076
Share of urban	46.785	24.317
Exports	34.250	24.308
Temperature	18.366	8.387

Table 8. Correlation matrix								
	Conflict	GDP Capita	Lag of GDP Capita	Share of young	Share of old	Share of urban	Exports	Temperature
Conflict	1							
GDP Capita	-0.1673	1						
Lag of GDP Capita	-0.1637	0.9992	1					
Share of young	0.1624	-0.7771	-0.7769	1				
Share of old	-0.155	0.7272	0.7271	-0.9136	1			
Share of urban	-0.0949	0.8268	0.8286	-0.6954	0.6238	1		
Exports	-0.1862	0.3218	0.3208	-0.2329	0.1182	0.303	1	
Temperature	0.1268	-0.5136	-0.5138	0.66	-0.7127	-0.4603	0.0315	1

Table 9. The Effect of Income on Conflict

	(1)	(2)	(3)	(4)	(5)
GDP capita	-0.0351***	-0.629***	-3.792***	-0.520***	-0.0880***
	(0.00327)	(0.0847)	(0.554)	(0.0896)	(0.0160)
Lag of GDP capita		0.597***	3.569***	0.486***	
		(0.0850)	(0.556)	(0.0903)	
Share of young				0.000162	
				(0.00153)	
Share of old				-0.00582	
				(0.00360)	
Share of urban				0.00248***	
				(0.000409)	
Exports				-0.00301***	
				(0.000267)	
Temperature				0.00393***	
				(0.000984)	
Constant	0.445***	0.434***	0.174	0.371***	0.838***
	(0.0249)	(0.0251)	(0.171)	(0.0950)	(0.119)
Observations	5,532	5,415	5,415	4,635	5,532
R-squared	0.020	0.029		0.071	0.418
Fixed Effects	no	no		no	yes
Estimator	OLS	OLS	Logit	OLS	Fixed Effects
F-statistic	115.3	80.46		50.35	30.12

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable in all regressions is an indicator variable for Conflict occurrence.