Økonometri A
Efterår 2024

Mandatory Assignment 3

US Jobs and Import Competition from China

This assignment is the third of three mandatory assignments in Econometrics A. All three assignments must be passed in order to go to the final exam.

The assignment may be answered in groups of max 3 students.

This assignment is the same as the final exam in Econometrics 1 (Polit) in the Summer 2017.

Practical instructions to the assignment
Read the entire assignment before you begin to respond, and answer all the questions.

The answer to the assignment must be presented in a comprehensive report with relevant tables and figures. The front page of the report must be based on the template that is available on Absalon. Fill in the names and study ID of all group members on the front page.

Prepare one Python Jupyter Notebook .ipynb-file that generates all tables and figures that appear in your report. The program should produce tables and figures in the same order as they appear in the report. Comments should clearly indicate which table or figure appearing in the report is being produced. Make sure that the .ipynb-file can be executed without any errors. The .ipynb-file must include the names and study ID of all group members.

To pass a mandatory assignment, it is required that:

- An adequate response is given to all questions in the assignment.
- More than half of the questions are answered correctly.
- Answers are written in a precise and easy to understand language.

The report must not exceed 8 A4 pages with font size 12, line space 1.5 and page margins of 2.5 cm. This includes the main text, tables and figures in the report, but not the front page.

Uploading your report
Each group must hand-in only one report in total. You must upload two files:

1. The report itself must be uploaded as a PDF file.
2. The Python Jupyter Notebook .ipynb-file must be uploaded as a file in plain text format (.txt).

If needed, a free PDF converter can be found here: www.pdf995.com

If group members are assigned to different class teachers, upload the report to the class teacher of your choice. Do not upload the same report to more than ones.

Introduction to the assignment:

"US Jobs and Import Competition from China"

A main theme of the recent US presidential election is the issue of globalization and its effects on American workers. During the election campaign, candidate Trump often blamed China for the loss of manufacturing jobs in the US, and now, part of his "Make America Great Again" plan is to bring back domestic jobs by re-negotiating existing international trade agreements.

Economists have long recognized a link between domestic jobs and international trade. One basic observation of the US economy is that the fraction of workers employed in the manufacturing sector fell by a third between 1990 and 2007. In the same period, US imports from low-wage countries increased from 9 to 28 percent of total imports, with China accounting for 89 percent of this growth. Much of this change has been attributed to the economic miracle of China which is largely due to its transition to a market-oriented economy as well as China's accession to the World Trade Organization (WTO) in 2001. One would expect that lower trade barriers between the US and China cause some US jobs to move to China where they can be performed at lower costs. With production increasingly taking place in China, final goods are then imported to the US for final consumption instead of being produced locally. This is known as import competition.

Autor, Dorn and Hanson (2014) present an analysis of the effect of rising Chinese import competition on US manufacturing jobs.[2] Their analysis is based on data on a larger number of geographical areas in the US that differ in the relative importance of different manufacturing industries for local employment – and their exposure to import competition from China.

A key challenge when estimating the employment effects of import competition is the issue of confounding factors. For instance, the substitution of US workers for Chinese workers in manufacturing production suggests that import competition and employment are negatively related. On the other hand, unobserved local demand shocks may increase overall labor demand for both US and Chinese workers, leading to a positive relationship between import competition and employment. To isolate the impact of Chinese import competition, Autor, Dorn and Hanson argue that the global rise in imports from China reflects an exogenous change in China's global competitiveness that is common to all countries in the world. Moreover, exogenous changes in China's supply conditions have had a different impact on different geographical areas in the US, depending on the initial patterns of industry specialization across regions in the US.

In this exam, you are asked to estimate the causal effect of Chinese import competition on US manufacturing jobs.

[2]This exam is inspired by David Autor, David Dorn and Gordon Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States", *American Economic Review*, 2014, pp. 2121-2168. The data used in this exam are simulated, and it is not possible to replicate the results in the paper using the data provided here.

Documentation of the data

The data consists of 1,324 observations on 662 commuting zones in the US for two time periods, 1990-2000 and 2000-2007. Commuting zones are defined as geographical areas with strong economic relationships within a given region of the US. As such, each commuting zone may be thought of as geographical subeconomy of the aggregate US economy. The following variables are available for analysis:

Table 1: List of variables

STATA name	Text label	Description
czone	c	Commuting zone identifier.
t2	$t2$	Time dummy, indicating if an observation belongs to the second period (2000-2007) or not.
dsL	ΔsL_{ct}^{man}	Percentage change in the manufacturing employment share in commuting zone c in period t.
dIPWusch	ΔIPW_{ct}^{USCH}	Percentage change in US imports from China per worker in commuting zone c in period t.
dIPWusmx	ΔIPW_{ct}^{USMX}	Percentage change in US imports from Mexico per worker in commuting zone c in period t.
college	$college_{ct}$	Percentage of college-educated population in commuting zone c in the first year of period t.
foreignborn	$foreignborn_{ct}$	Percentage of foreign-born population in commuting zone c in the first year of period t.
routine	$routine_{ct}$	Percentage of employment in routine-intensive jobs in commuting zone c in the first year of period t.
dIPWotch	ΔIPW_{ct}^{OTCH}	Percentage change in other countries imports from China per worker in commuting zone c in period t.
dIPWukch	ΔIPW_{ct}^{UKCH}	Percentage change in UK imports from China per worker in commuting zone c in period t.
dIPWotmx	ΔIPW_{ct}^{OTMX}	Percentage change in other countries imports from Mexico per worker in commuting zone c in period t.

Note 1: All variables expressed in percentage changes represent annual average changes within the two periods. In this way, percentage changes across the two time periods can be compared even though the periods are of different lengths.

Note 2: Other countries include Australia, Finland, Germany, Japan and the UK.

Note 3: The variables ΔIPW_{ct}^{OTCH}, ΔIPW_{ct}^{UKCH} and ΔIPW_{ct}^{OTMX} are defined in similar ways. To clarify how they are constructed, let's take a look at ΔIPW_{ct}^{UKCH}. Suppose San Francisco is a commuting zone in the data. At the beginning of a period t, San Francisco has employment in various industries (say, agriculture, IT and many others). Imagine this employment for each industry is expressed relative to the US aggregate employment in every industry (say, San Francisco employs 10 percent of all IT workers in the US, but only 0.01 percent of all agricultural workers). ΔIPW_{ct}^{UKCH} is calculated by multiplying these initial employment shares with the changes in UK imports per worker for each industry and then summing up across industries. One way to think about this is: Imagine San Francisco is a commuting zone in the UK instead of the US, but with the same employment characteristics. What is the hypothetical change in San Francisco imports from China in this scenario? ΔIPW_{ct}^{UKCH} for $c =$ San Francisco!

Problem 1 (20%)

1. Provide a descriptive analysis of the variables in your data using relevant summary statistics. Describe how US imports have changed across commuting zones between 1990 and 2007.

2. Consider the regression model:

$$\Delta sL_{ct}^{man} = \beta_0 + \delta_0 t2_t + \beta_1 \Delta IPW_{ct}^{USCH} + \delta X_{ct} + u_{ct} \tag{1}$$

   where X_{ct} includes $college_{ct}$, $foreignborn_{ct}$ and $routine_{ct}$.

   (a) What is the interpretation of β_1?

   (b) What is the expected sign of β_1?

   (c) Estimate the parameters of model (1) using OLS. Report your estimates in a table with relevant standard errors. Is the OLS estimate of β_1 consistent with your expectation?

Problem 2 (20%)

Unobserved local demand shocks may affect the estimated relationship between the employment share of US manufacturing and import competition from China. Autor, Dorn and Hanson (2014) argue that rising Chinese import competition reflects mostly changes in the supply conditions of producers in China. They argue that changes in China's production supply in recent decades have led to rising imports from China to the US as well as to many other countries in the world. As such, the authors view the global increase in Chinese import competition as a result of exogenous changes in China's global competitiveness.

1. Discuss the conditions needed for ΔIPW_{ct}^{OTCH} to be a relevant and valid instrument for ΔIPW_{ct}^{USCH} in model (1). Are the conditions likely to be satisfied in this case? Present empirical evidence as needed to support your answer.

2. Estimate the parameters of model (1) using IV. Report your results in a table and discuss how they compare to your OLS results from Problem 1.

3. ΔIPW_{ct}^{OTCH} is based on the imports from China to a group of other countries. The UK is part of this group of other countries, but a separate measure of UK imports from China is available. Use ΔIPW_{ct}^{UKCH} as an additional IV and implement a test of overidentifying restrictions.

4. US manufacturing jobs may also relocate to Mexico due to lower Mexican production costs. Extend model (1) with ΔIPW_{ct}^{USMX} and estimate the parameters by IV. Is the impact of Mexican and Chinese import competition on US manufacturing different?

Hint for 2.4:

When estimating 2SLS models using the package linearmodels in Python, you can specify several endogenous variables and several instruments in your model formula like this:
[dIPWusch + dIPWusmx ~ dIPWotch + dIPWothmx]
Here, dIPWusch and dIPWusmx are endogenous, while dIPWoth and dIPWothmx serve as instruments.

Problem 3 (20%)

US imports from China have increased dramatically in recent decades. Some commentators argue that the impact of Chinese import competition itself has changed as well. To address this issue, consider an extended version of model (1):

$$\Delta sL_{ct}^{man} = \beta_0 + \delta_0 t2_t + \beta_1 \Delta IPW_{ct}^{USCH} + \beta_2 \left(t2_t \times \Delta IPW_{ct}^{USCH}\right) + \delta X_{ct} + u_{ct} \qquad (2)$$

1. Estimate the parameters of model (2) using 2SLS with ΔIPW_{ct}^{OTCH} and $(t2_t \times \Delta IPW_{ct}^{OTCH})$ as IVs. Implement the 2SLS estimator manually. Report the first and second stage results in a table and comment on your results.

2. Estimate the parameters of model (2) using IV and test for no difference in the impact of import competition across the two time periods.

3. Conduct a test for exogeneity of ΔIPW^{USCH} and $t2 \times \Delta IPW^{USCH}$.

Hint for 3.1:
Here, the intention is that you run the two first stage regressions and the second stage regression manually. You can for example use statsmodels for this. You don't have to manually code up the 2SLS estimator.

Problem 4 (20%)

Measurement error in an explanatory variable may lead to attenuation bias in the OLS estimator. This problem may be corrected for using a relevant and valid instrument. That said, one concern is that the instrument itself is measured with error which may create an additional source of bias. To address this issue consider a simple regression model:

$$y_i = \beta_0 + \beta_1 x_i^* + u_i \qquad (3)$$

where x_i^* is the true but unobserved explanatory variable. The observed explanatory variable, x_i, is defined as $x_i = x_i^* + \epsilon_i$, where ϵ_i is a measurement error that is uncorrelated with the true explanatory variable, i.e., $cov(x^*, \epsilon) = 0$. An instrument, z, is available, but it is measured with error. Assume $z_i = z_i^* + \eta_i$, where z_i^* is the true but unobserved instrument and η is a measurement error. The true instrument is related to the true explanatory variable, $cov(z^*, x^*) = \theta\sigma_{x^*}^2$, where θ is a parameter and $\sigma_{x^*}^2$ is the variance of x^*. The measurement errors in the two variables may be correlated, $cov(\epsilon, \eta) = \rho\sigma_\epsilon^2$, where ρ is a parameter and σ_ϵ^2 is the variance of ϵ. To simplify the analysis below, assume that u is independent of z as well as $cov(z^*, \epsilon) = cov(x^*, \eta) = 0$.

1. Derive the probability limit of the IV estimator of β_1 under these assumptions. For what values of ρ is the IV estimator a consistent estimator of β_1?

2. For what values of θ and ρ are the probability limits of the OLS and IV estimators the same, i.e., $plim\, \widehat{\beta}_1^{OLS} = plim\, \widehat{\beta}_1^{IV}$?

Problem 5 (20%)

Consider the following data generating process (DGP):

$$y_i = \beta_0 + \beta_1 x_i^* + u_i \tag{4}$$

$$x_i = x_i^* + \epsilon_i \tag{5}$$

$$z_i = \theta x_i^* + \eta_i \tag{6}$$

$$\eta_i = \rho \epsilon_i + \mu_i \tag{7}$$

$$\beta_0 = 4,\ \beta_1 = 3,\ \theta = 1 \tag{8}$$

$$x^* \sim N(1,4),\ u \sim N(0,1), \epsilon \sim N(0,1),\ \mu \sim N(0,1) \tag{9}$$

Implement a simulation experiment of $\widehat{\beta}_1^{OLS}$ and $\widehat{\beta}_1^{IV}$ based on the above DGP, where y is the dependent variable, x is the observed explanatory variable, and z is the observed instrument. Consider the following values of ρ: -0.5, 0, 0.5 and 1. Pick a seed number of your choice and write it explicitly in the main text of your report. Draw samples of 1,000 observations and replicate the experiment 500 times. Report summary statistics and histograms to document the results from the simulation experiment. Are the simulation results in line with the analytical statements from Problem 4? Discuss and compare your results. Does the IV estimator suffer from attenuation bias in the presence of measurement errors?