Econometrics Final Project

Theodore Peterson, Graham Arthur, Dorsa Massihpour

Contents

Introduction

"The China Syndrome: Local Labor Market Effects of Import Competition in the United States" by Autor, Dorn, and Hansen seeks to analyze the effect of rising Chinese import competition on U.S. manufacturing jobs. The study claims that a quarter of the decline in U.S. manufacturing employment between 1990 and 2007 is explained by increased import competition from China.

Due to the complex nature of this problem, several specifications were used in order to obtain a causal effect. Some methodologies, such as panel data and instrumental variables analysis were used to negate the pitfalls of estimating an effect this complex.

For this analysis, the authors observed 722 regions of the United States, given by Commuter Zones (CZs) developed by Tolbert and Sizer (2006). These zones represent the local labor markets on which the study examined the effects of Chinese import competition.

Variables and Control Measures

The outcome variable, employment, will be measured by ΔL_{it}^{m}, the decadal change in manufacturing employment share in commuter zone i. In simpler terms, this describes the mixing of employment between manufacturing and non-manufacturing sectors within a given region, over a decade.

Our primary predictor variable is ΔIPW_{uit}, the change in Chinese import exposure per capita in commuter zone i. This measure is constructed by multiplying the change US imports from China per region capita $(\Delta M_{ucjt}/L_{it})$ by the region's share of national employment (L_{ijt}/L_{ujt}). This metric is then aggregated across industries (j) to obtain the final measure.

$$\Delta IPW_{uit} = \sum_{j} (\frac{L_{ijt}}{L_{ujt}} \times \frac{\Delta M_{ucjt}}{L_{it}})$$

The predictor variable, ΔIPW_{uit} will be instrumented using Chinese import exposure to eight other high income markets ΔIPW_{oit}. This measure is constructed similarly to the above, except that instead of Chinese imports to the US (ΔM_{ucjt}), we use Chinese imports to those other markets (ΔM_{ocjt}). The measure also eliminates simultineity bias by using lagging employment measures.

$$\Delta IPW_{oit} = \sum_{j} (\frac{L_{ij(t-1)}}{L_{uj(t-1)}} \times \frac{\Delta M_{ocjt}}{L_{i(t-1)}})$$

The main control described is the start-of-period manufacturing employment share. According to the authors, this is included to focus on the effect of industry mix within local manufacturing sectors. The paper also includes a "rich set" of controls for each commuter zone, including demographic and labor force measurements.

Opinion

In our opinion, the variables suggested do provide an adequate measure of the concepts that we are trying to capture. The extra controls mentioned are particularly appeasing since it is likely that regions will differ in ways that could affect both employment and their imports from China. In addition, we believe that the instrumental variable and panel data approaches will remove many of our inital concerns about the model.

Summary Statistics

We were able to recreate all the means from Appendix Table 2 and the import/export growth percentage changes from Table 1. For the standard deviations in Appendix Table 2 are normalized, so our results are the same when normalization is accounted for. This is only due to differences between Stata and R. In order to obtain the summary statistics table, read the instructions at the end of the code appendix in the RMD file.

Specification

Structural Equations and Assumptions

The structural equation is given by:

$$L_{it}^{m} = \gamma_t + \beta_1 IPW_{uit} + \beta_2 X_{it}' + a_i + e_{it}$$

The assumptions under which $\hat{\beta}_1$ is an unbiased and consistent estimator are as follows:

1. There is a true linear model in the population.

2. The units of observation are randomly sampled from a population. (since the U.S. is the population, the 722 commuter zones effectively represent it)
3. There are no exact linear relationships among the predictor variables. This includes relationships between IPW_{uit} and elements of X'_{it}, as well as between elements of X'_{it}.
4. There is no correlation between the error terms e_{it} and any of the predictors. (this assumption is violated, hence the IV approach)

Note that because we are using a first differences model, our equation will become:

$$\Delta L^m_{it} = \gamma_t + \beta_1 \Delta IPW_{uit} + \beta_2 X'_{it} + e_{it}$$

Residuals and Unobserved Heterogeneity

In the model given above, the residuals $\hat{e}_{it}$ estimate the percentage change in a region's manufacturing job share coming from unobserved shocks to the economies of either China or the US.

There exist some sources of heterogeneity that will likely be captured by X', the set of controls for each region which includes both demographic and labor market statistics. However, the authors mention unobserved heterogeneity from three main sources:

- Start-of-period manufacturing employment share. This is a component of X' in later specifications, so it does not pose a significant threat.
- The share of trade imbalance in total expenditure. This variable, denoted by ρ_i, represents the share of the current account deficit in total expenditure in region i.
- The equilibrium scaling factor. This variable, c_i is simply a constant used to correct measures of log changes in employment that provide the motivations for the model.

The authors control for start-of-period manufacturing share, but they assume that ρ_i, c_i are identical for all regions such that $\rho_{ij} c_i = \alpha$. This may make empirical analysis easier, however it also assumes homogeneity among regions. This is not accounted for in the analysis, so it is something to remember.

Sources of Endogeneity

The authors describe several potential sources of endogeneity that may be corrected with the instrument, and some that may persist even after instrumentation.

- Unobserved shocks to the US economy could affect both region employment and imports per capita. This would imply a correlation between the error terms and our predictor. Most of this effect is mitigated by using an instrumented approach, since we expect shocks to the US economy to have less of an effect on the imports of other nations.
- Despite instrumentation, the authors admit that there still may be correlation between the imports of other nations and unobserved shocks. One of these is the potential correlation between global economic shocks that effect all of the high-income markets we are studying, including the US.
- Another threat after instrumentation is the possibility that US shocks affect global imports. The example given describes a failure in US markets that provides an opportunity for China to export more to both the US and other high-income markets. This again produces a correlation between our error terms and our instrument.

- The last threat to instrumentation that is mentioned is the similarity of the high-income markets with the US. The paper mentions that technological shocks are common to these markets, and that they can harm manufacturing sectors in those countries, which again provides correlation between the error term and the instrument.

For these endogeneity threats, if there is sufficient insulation between any of the global economic shocks mentioned (e_{it}) and the per capita import exposure of high-income markets (ΔIPW_{oit}), then the instrumental variables approach will give an unbiased and consistent estimator for the β_1 described above.

Replication of Main Analysis

We were able to accurately recreate the results in Tables 2 and 3, weighting the data by the appropriate population, and clustering by 'statefip' to get the appropriate standard errors. We estimate that a $1,000 increase in import exposure per worker will generate a 0.75% decrease in manufacturing employment in our unspecified model, and a 0.60% decrease in our final specification. These effects (e.g. decreased employment) were statistically significant.

2SLS Estimate Validity

After replicating the data from Table 3, we found that all the first stage estimates were significant, just like the study did. We also performed the F test and found the instruments were not weak (rejected null hypothesis at $\alpha \approx 0$ level). We also performed the Hausman test and found that we do have endogeneity the IV regression was indeed efficient. We rejected the null hypothesis that OLS and IV estimates are the same at the $\alpha \approx 0$ level. The Sargan test for overidentification was not applicable here because the model only had one instrument and therefore was not overidentified. It was exactly identified 1:1.

The OLS estimates are smaller than the than the 2SLS estimates, as was predicted by the authors. This means that for the purposes of isolating supply shocks to U.S. producers, the instrumental variables do the job better than a simple OLS. By all indications from the tests we performed, the authors seem to be correct in approaching their regression from an instrumental variables strategy and have chosen an instrument that will fulfill the intended purpose. They are relevant and exogenous.

Analysis

Misspecification

The instrumental variables approach removes much of the endogeneity present in the original model. Despite this exercise, the authors have failed to guard against many of the threats mentioned above. For example, a supply shock in the U.S. could occur at the same time as supply shocks in the eight other developed nations in our IV. This may occur because of events affecting the worldwide economy, such as natural disasters influencing multinational companies in the agricultural or natural resources industries. Since developed nations may be more adept at getting through these disasters, other countries may rely on their goods. Additionally, there may be demand shocks that occur simultaneously in the U.S. and these other countries, possibly because of growing populations due to better living conditions.

In order for the exclusion principle to be satisfied, growth of Chinese import exposure to other high-income markets (ΔIPW_{oit}, our IV) must only directly affect the U.S. labor market's exposure to import competition (exogeneity condition). The labor market's exposure to import competition (ΔIPW_{uit}) must in turn affects the decadal change in the manufacturing employment share of the working-age population in region i (relevance condition). The exogeneity condition should be satisfied because the Chinese import of exposure of other developed nations should not be correlated to the change in the manufacturing employment rate in the

U.S. (unless the sources of endogeneity mentioned above occur). Intuitively, other countries' import exposure should only affect employment rates within their own countries. Another exception to this may occur if there is competition between developed countries in importing China's goods. For example, if the U.S. needs to import a Chinese good but cannot because another developed country is importing most of China's supply, the U.S. employment rate may increase to make up for the amount of that good needed in the U.S. To the extent that none of these issues occur, the exclusion principle is satisfied.

There also may be functional forms that are misspecified. Trending is of no particular concern since both our predictor and instrument ($\Delta IPW_{uit}, \Delta IPW_{oit}$) are differenced. This should remove any stochastic or linear trends. Another avenue to consider is the addition of interaction variables. Specifically, issues may arise without including an interation term between the time dummy. For example, due to a policy shock, the ability of employment to be offshored in a region in 2007 may have a larger effect on employment than in 2000 or 1990. If we exclude these time dummy interactions, we may be oversimplifying the nature of the relationship, and our estimates would be subject to ommitted variables bias.

Alternative Specifications

- To account for import demand shocks in industries that may be correlated among CZ's, separate OLS and 2SLS estimates were calculated with specific industries dropped.
- Adding a set of demographic and labor force measures to account for confounding factors. Within this specification, a control was added for the share of manufacturing in a CZ's start period employment (in order to make sure that increase in import exposure is not due to an underlying trend in the U.S. and is actually due to exposure imports causing differences in manufacturing industries across CZ's). Using this specification, the effect of import exposure on manufacturing employment is slightly smaller than that in the original specification, but is still statistically significant.
- Adding geographic dummies for the 9 Census divisions (from which data for this paper was taken) to the original model in order to absorb region-specific trends in manufacturing employment. This specification also included controls for start of period population ratio that had a college education (across CZ's), share of population that is foreign born, and share of working-age women who are employed. This specification showed a decreased effect of import exposure on manufacturing employment.
- Adding variables that account for the vulnerability of a CZ's occupations to be taken by technology (routine share variable) or task offshoring. One main category of occupations was routine-intensive occupations that are often replaced with computers. CZ's which have a high level of these types of jobs were found to have a strong positive correlation between the routine share variable and change in manufacturing share. However, the offshoring index, which measure the extent to which jobs in CZ's require no face-to-face contact or proximity with workers in the U.S., did not have a statistically significant negative correlation with manufacturing employment share. Overall, including the routine share variable and offshoring index into the model showed a larger decline in manufacturing employment with a greater initial manufacturing employment share, and a smaller decline when the initial population contains more foreigners.

Potential Sources of Error

Misspecification also occurs because there is serial correlation in the exposure to imports. This is evident when testing that the causal relationship in the empirical model is true, and that other factors did not cause a decline in U.S. manufacturing even before import exposure occurred. The CZ's with the highest amounts of trade exposure in the 2000s vs. the 1990s were found to have their manufacturing employment negatively correlated with contemporaneous trade exposure, but approximately uncorrelated with future trade exposure. Since the negative relationship between manufacturing employment and contemporaneous trade exposure was statistically significant, serial correlation in exposure imports across CZ's can be concluded.

Conclusion

There are no clear violations of the assumptions of the model, however we still have doubts on the endogeneity assumptions that arise from our instrumental variables approach. The authors provide some rationale for this assumption, but little empirical evidence. To the extent that the instrument is not exogenous to the model, our estimate for the effect is inconsistent.

Another area of concern is the lack of interaction terms to describe the complex relationships between time and certain features of each region. If such an effect was significant, it would subject the model to omitted variable bias that would cause an inaccurate estimate.

Despite these caveats, we believe that the model does provide a credible estimate of the causal effect of Chinese import competition on manufacturing employment in the U.S. Anectodal evidence supports the exogeneity assumption, and the addition of more interaction terms will likely be a tedious exercise with little effect on the specification. When replicating the original procedure, we found that the increase in imports of Chinese goods over the time period studied has had negative effects on U.S. labor markets exposed to increased import competition. By combining an instrumental variables approach with panel data techniques, the authors have managed to circumnavigate many of the complications that would arise from a problem this complex.

Appendix

Table 1: Information on Trade with China

	1
percent_19912007_chus	1156
percent_19912007_usch	456
percent_19912007_uslw	491
percent_19912007_usce	375
percent_19912007_ushi	137
percent_19912007_chot	832
percent_19912007_otch	639
percent_19912007_otlw	236
percent_19912007_otce	316
percent_19912007_othi	84

Table 2: 2SLS Estimates

	Dependent variable:					
	(1)	(2)	(3)	(4)	(5)	(6)
d_tradeusch_pw	-0.888^{***}	-0.718^{***}	-0.746^{***}			
	(0.183)	(0.065)	(0.069)			
t2000			0.444			
			(0.327)			
d_tradeusch_pw_future				0.431^{***}	-0.130	0.148
				(0.151)	(0.127)	(0.096)
t1980						-1.945^{***}
						(0.250)
Constant	-1.056^{***}	-0.846^{***}	-1.218^{***}	-0.954^{***}	-1.832^{***}	-0.415
	(0.195)	(0.258)	(0.140)	(0.319)	(0.334)	(0.303)
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 3: Full Analysis

	Dependent variable:					
	(1)	(2)	(3)	(4)	(5)	(6)
d_tradeusch_pw	−0.746***	−0.610***	−0.538***	−0.508***	−0.562***	−0.596***
	(0.069)	(0.095)	(0.092)	(0.082)	(0.098)	(0.100)
l_shind_manuf_cbp		−0.035	−0.052**	−0.061***	−0.056***	−0.040***
		(0.022)	(0.020)	(0.017)	(0.017)	(0.013)
reg_midatl			0.182	−0.159	0.401**	0.313
			(0.202)	(0.199)	(0.171)	(0.286)
reg_encen			0.889***	0.663**	1.184***	1.261***
			(0.271)	(0.305)	(0.269)	(0.343)
reg_wncen			1.692***	1.681***	1.479***	1.624***
			(0.468)	(0.538)	(0.404)	(0.378)
reg_satl			−0.124	−0.437	−0.294	−0.288
			(0.272)	(0.290)	(0.257)	(0.237)
reg_escen			1.094***	0.451	0.932***	1.076***
			(0.279)	(0.301)	(0.233)	(0.340)
reg_wscen			1.134***	0.537**	0.797***	0.732***
			(0.165)	(0.222)	(0.156)	(0.235)
reg_mount			0.759***	0.449	0.398*	0.402
			(0.264)	(0.288)	(0.226)	(0.261)
reg_pacif			0.591***	0.274	0.480***	0.027
			(0.151)	(0.215)	(0.177)	(0.194)
l_sh_popedu_c				−0.008		0.013
				(0.017)		(0.012)
l_sh_popfborn				−0.007		0.030***
				(0.008)		(0.011)
l_sh_empl_f				−0.054**		−0.006
				(0.025)		(0.025)
l_sh_routine33					−0.230***	−0.245***
					(0.064)	(0.065)
l_task_outsource					0.244	−0.059
					(0.256)	(0.241)
t2	0.444	0.085	−0.096	−0.062	−0.119	−0.242
	(0.327)	(0.371)	(0.373)	(0.320)	(0.369)	(0.407)
Constant	−1.218***	−0.638*	−0.975***	3.407***	6.573***	6.279***
	(0.140)	(0.343)	(0.354)	(1.181)	(2.132)	(1.969)
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 4: (First Stage)

	Dependent variable:					
	(1)	(2)	(3)	(4)	(5)	(6)
d_tradeotch_pw_lag	0.792***	0.664***	0.652***	0.635***	0.638***	0.631***
	(0.080)	(0.089)	(0.094)	(0.094)	(0.091)	(0.091)
Note:						*p<0.1; **p<0.05; ***p<0.01