Written Exam for the B.Sc. or M.Sc. in Economics Summer 2017

Økonometri I/Econometrics I

Take-home exam

May 31, 2017

SUGGESTED ANSWERS

The assessment of the take-home exam is based on the handed-in report. The STATA program is not assessed as such, but is used e.g. to clear up errors and ambiguities in the report and to ensure that no collaboration across groups has taken place.

The assessment takes into account whether the report is overall correct and complete, and whether it, within the specified framework, is able to answer the problems stated in the assignment. The report should not exceed the total page numbers stated in the assignment. Should the page number exceed this, it will influence the overall grading negatively.

For more information about Econometrics I and the learning objectives of the course, please go to the course website:

`http://kurser.ku.dk/course/a%C3%98kb08020u/2016-2017`

The numerical results and conclusions in the suggested answers are based on GROUPDATA0.dta. Full results and conclusions for other data sets can be obtained by running the STATA program X2017S_takehome.do with the relevant data file. Results and conclusions may differ for different data sets.

Problem 1.1

Variables should as a minimum be described by a table indicating the average, standard deviation, min and max. Comments must address characteristics of key variables used in subsequent analyses.

The data include 1,324 observations on 662 commuting zones in the US. The average percentage change in the manufacturing employment share is -1.7 percent which is consistent with the aggregate loss of US manufacturing jobs in the period 1990-2007. That said, the decline of US manufacturing is very unevenly dispersed across different commuting zones as seen from the minimum and maximum observations. In contrast, US imports from China have increased by an average of 2 percent, which is more than double the increase of US imports from Mexico. The negative change in the employment share of US manufacturing and the positive change in US imports from China are particularly noticeable in the period between 2000 and 2007.

Table 1: Summary Statistics, 1990-2007

	mean	p50	min	max
dsL	-1.65	-1.38	-19.17	5.66
dIPWusch	1.93	1.18	-0.63	43.08
dIPWusmx	0.91	0.43	-6.34	12.60
college	45.16	45.30	19.94	69.34
foreignborn	4.97	2.94	0.38	48.91
routine	28.54	28.60	19.99	37.75
N	1324			

Table 2: Summary Statistics, 1990-2000

	mean	p50	min	max
dsL	-0.96	-0.69	-19.17	5.66
dIPWusch	1.19	0.74	-0.08	25.41
dIPWusmx	1.39	0.82	-0.47	12.60
college	42.10	42.52	19.94	62.90
foreignborn	3.92	2.18	0.38	39.94
routine	28.29	28.26	19.99	37.75
N	662			

Table 3: Summary Statistics, 2000-2007

	mean	p50	min	max
dsL	-2.34	-2.07	-14.38	4.64
dIPWusch	2.68	1.94	-0.63	43.08
dIPWusmx	0.44	0.23	-6.34	7.51
college	48.23	49.11	26.32	69.34
foreignborn	6.03	3.77	0.73	48.91
routine	28.78	28.84	22.23	36.66
N	662			

Problem 1.2

(a) β_1 describes the expected percentage change in the employment share of US manufacturing from increasing US imports from China per worker with one percentage point (holding the other regressors constant).

(b) Import competition is often considered to be a substitute for domestic production, suggesting $\beta_1 < 0$.

(c) The OLS results are presented in column 1 in Table 4. The estimated $\widehat{\beta}_1$ indicates that a one-percentage point higher import competition from China lowers the manufacturing employment share in the US by -0.154 percentage points. The estimate is significantly different from zero at the 1-percent level. This finding is consistent with job losses in US manufacturing being negatively related to rising Chinese import competition.

Table 4: OLS and IV results

	(1) OLS (Prob 1.2)	(2) IV (Prob 2.2)	(3) IV (Prob 2.3)	(4) IV (Prob 2.4)
t2	-1.509*	-1.082*	-1.133*	-1.358*
	(0.128)	(0.202)	(0.174)	(0.374)
dIPWusch	-0.154*	-0.380*	-0.353*	-0.340*
	(0.0375)	(0.100)	(0.0786)	(0.101)
college	0.0836*	0.0687*	0.0704*	0.0606*
	(0.00737)	(0.00951)	(0.00830)	(0.0131)
foreignborn	0.00460	-0.00714	-0.00573	-0.0122
	(0.00671)	(0.00768)	(0.00720)	(0.0100)
routine	-0.333*	-0.280*	-0.286*	-0.251*
	(0.0219)	(0.0305)	(0.0258)	(0.0437)
dIPWusmx				-0.277
				(0.323)
_cons	5.098*	4.543*	4.609*	4.430*
	(0.491)	(0.541)	(0.512)	(0.549)
N	1324	1324	1324	1324

Standard errors in parentheses

* $p < 0.05$

Problem 2.1

The instrument, ΔIPW^{OTCH}, is relevant if it is correlated with the endogenous variable, ΔIPW^{USCH}. This may be tested using the first stage regression:

$$\Delta IPW_{ct}^{USCH} = \pi_0 + \pi_1 t2_t + \pi_2 \Delta IPW_{ct}^{OTCH} + \Pi X_{ct} + e_{ct}$$

If rising Chinese import competition in the US reflects increases in China's global competitiveness, one would expect $\pi_2 \neq 0$. The 1st stage results are presented in column 1 of Table 5. The null hypothesis of no instrument relevance is $H_0 : \pi_2 = 0$ and the alternative is $H_A : \pi_2 \neq 0$. The test statistic is $t = \frac{\widehat{\pi}_2}{HCSE(\widehat{\pi}_2)} = \frac{0.827}{0.156} = 5.3$. This is larger than the critical value of 1.96 (the 97.5th percentile of a standard normal) and the null hypothesis is rejected in favor of the alternative hypothesis using a 5 percent significance level. This evidence suggests that rising Chinese import competition in other countries can explain rising Chinese import competition in the US. The instrument is valid if Chinese import competition in other countries is unrelated to unobserved local demand shocks in the US, i.e., $cov(\Delta IPW^{OTCH}, u) = 0$. As such, the instrument is only valid insofar unobserved local demand shocks are uncorrelated across the US and other countries in the world. Only in this scenario will the instrument capture exogenous changes in China's supply conditions.

Table 5: First stage regressions (Problem 2)

	(1) dIPWusch	(2) dIPWusch	(3) dIPWusch	(4) dIPWusmx
t2	0.303	-0.266*	0.338	-1.016*
	(0.278)	(0.0493)	(0.265)	(0.0933)
dIPWotch	0.827*	0.0741*	0.834*	0.105*
	(0.156)	(0.0145)	(0.159)	(0.0314)
college	-0.00788	0.0158*	-0.00803	-0.0299*
	(0.0118)	(0.00257)	(0.0118)	(0.00495)
foreignborn	-0.0191*	-0.00164	-0.0199*	-0.0195*
	(0.00576)	(0.00359)	(0.00558)	(0.00584)
routine	0.0659*	-0.0641*	0.0707*	0.104*
	(0.0335)	(0.00707)	(0.0318)	(0.0138)
dIPWukch		1.654*		
		(0.0197)		
dIPWotmx			-0.670	1.230*
			(0.489)	(0.391)
_cons	-1.111*	1.062*	-1.235*	-0.338
	(0.453)	(0.203)	(0.431)	(0.387)
N	1324	1324	1324	1324

t statistics in parentheses

* $p < 0.05$

Problem 2.2

The IV estimates are presented in column 2 of Table 4. The effect of rising Chinese import competition is estimated to be -0.38 according to the IV estimates, which is more than twice the magnitude of the OLS estimate. This finding is consistent with the view that the OLS estimate is upward biased due to unobserved local demand shocks that introduce a positive correlation between import competition and employment shares. The standard error is higher on the IV estimate compared to the OLS estimate, but the estimated effect remains statistically significant.

Problem 2.3

The new instrument, ΔIPW^{UKCH}, is part of the instrument from the previous problem. As such, both instruments are trying to capture exogenous variation in China's supply conditions. Consider the first stage regression:

$$\Delta IPW_{ct}^{USCH} = \pi_0 + \pi_1 t2_t + \pi_2 \Delta IPW_{ct}^{OTCH} + \pi_3 \Delta IPW_{ct}^{UKCH} + \Pi X_{ct} + e_{ct}$$

The first stage estimates are presented in column 2 of Table 5. The null hypothesis of no instrument relevance is $H_0 : \pi_2 = 0,\ \pi_3 = 0$ is strongly rejected with an F statistics of 6128 (the 95th percentile of an F distribution with $df = (2, \infty)$ is 3). Assuming no correlation among unobserved demand shocks across the world, the instruments are also valid.

The IV estimates are presented in column 3 of Table 4. The effect of Chinese import competition is estimated to be -0.35 which is similar to the IV estimate using only one instrument. From the IV regression, the residuals $\widehat{u}_{ct}$ are calculated. Consider the regression:

$$\widehat{u}_{ct} = \gamma_0 + \gamma_1 t2_t + \gamma_2 \Delta IPW_{ct}^{OTCH} + \gamma_3 \Delta IPW_{ct}^{UKCH} + \Gamma X_{ct} + \eta_{ct}$$

The test for overidentifying restrictions is calculated as $OI = n \times R^2 = 1324 \times 0.0007 = 0.88$. This is compared to the 95th percentile of a χ^2-distribution with one degree of freedom which is equal to 3.84. The null of no relationship between the IV residuals and the exogenous variables cannot be rejected. As such, the conclusion is that at least one of the instruments are uncorrelated with u.

Problem 2.4

The extended model includes Mexican import competition as an additional regressor:

$$\Delta sL_{ct}^{man} = \beta_0 + \delta_0 t2_t + \beta_1 \Delta IPW_{ct}^{USCH} + \beta_2 \Delta IPW_{ct}^{USMX} + \delta X_{ct} + u_{ct} \tag{1}$$

This regression model has two endogenous variables, ΔIPW^{USCH} and ΔIPW^{USMX} that may be correlated with unobserved local demand shocks. Two IV candidates are ΔIPW^{OTCH} and ΔIPW^{OTMX} that seek to capture exogenous changes in Chinese and Mexican import competition. The first stage results are presented in columns 3 and 4 of Table 5, where it can be seen that both IVs are significantly related to the endogenous variables. The IV results are presented in column 4 of Table 4. The estimated effect of Chinese import competition is -0.34 which is similar to previous IV estimates. Moreover, it is statistically significant. The estimated effect of Mexican import competition is -0.27, but we cannot reject that β_2 is zero in the population. As a result, import competition from China and Mexico have a different effect on the manufacturing employment share in the US (i.e., $H_0 : \beta_1 = \beta_2$ is rejected in favor of $H_A : \beta_1 \neq \beta_2$). This finding suggests a greater substitution between Chinese and American workers compared to Mexican and American workers.

Problem 3.1

The first stage regressions are:

$$\Delta IPW_{ct}^{USCH} = \pi_0^1 + \pi_1^1 t2_t + \pi_2^1 \Delta IPW_{ct}^{OTCH} + \pi_3^1 (t2_t \times \Delta IPW_{ct}^{OTCH}) + \Pi^1 X_{ct} + e_{ct}^1$$

$$\left(t2_t \times \Delta IPW_{ct}^{USCH}\right) = \pi_0^2 + \pi_1^2 t2_t + \pi_2^2 \Delta IPW_{ct}^{OTCH} + \pi_3^2 (t2_t \times \Delta IPW_{ct}^{OTCH}) + \Pi^2 X_{ct} + e_{ct}^2$$

The second stage regression is:

$$\Delta sL_{ct}^{man} = \pi_0 + \pi_1 t2_t + \pi_2 \widehat{\Delta IPW}_{ct}^{USCH} + \pi_3 (t2_t \times \widehat{\Delta IPW}_{ct}^{USCH}) + \Pi X_{ct} + u_{ct}$$

where $\hat{}$ denotes the predicted endogenous variables from the first stage regressions (based on OLS). The first and second stage regression results are presented in columns 1-3 in Table 6. The 2SLS estimates suggest that Chinese import competition has a negative effect of -0.27 on the manufacturing employment share in the first period. In the second period, the effect is -0.27+(-0.13)=-0.40. This finding supports the view that the negative effect of Chinese import competition has become stronger in recent years. That said, the 2SLS standard errors cannot be used for hypothesis testing since they are not valid.

Table 6: 2SLS/IV results (Problem 3)

	(1) 1st	(2) 1st	(3) 2nd	(4) IV
dIPWotch	0.914*	-0.0725		
	(0.214)	(0.0549)		
t2dIPWotch	-0.0978	0.896*		
	(0.229)	(0.117)		
t2	0.409	0.579	-0.922*	-0.922*
	(0.399)	(0.378)	(0.234)	(0.298)
college	-0.00604	-0.00822	0.0706*	0.0706*
	(0.0111)	(0.00954)	(0.00806)	(0.00906)
foreignborn	-0.0188*	-0.0110*	-0.00611	-0.00611
	(0.00559)	(0.00525)	(0.00695)	(0.00752)
routine	0.0621	0.0466	-0.285*	-0.285*
	(0.0322)	(0.0283)	(0.0254)	(0.0294)
dIPWuschhat			-0.270*	
			(0.102)	
t2dIPWuschhat			-0.125	
			(0.112)	
dIPWusch				-0.270*
				(0.120)
t2dIPWusch				-0.125
				(0.135)
_cons	-1.173*	-0.857*	4.486*	4.486*
	(0.475)	(0.426)	(0.513)	(0.549)
N	1324	1324	1324	1324

t statistics in parentheses

* $p < 0.05$

Problem 3.2

The IV results are presented in column 4 of Table 6. As expected, the 2SLS and the IV estimates are numerically identical. The null hypothesis of no difference in the impact of Chinese import competition is $H_0 : \beta_2 = 0$. The t statistic is calculated to be -0.92 with a p-value above 0.05. The null hypothesis cannot be rejected and it is concluded that the effect of Chinese import competition on US manufacturing is the same for both 1990-2000 and 2000-07.

Problem 3.3

The test for exogeneity is performed by running the regression:

$$\begin{aligned}\Delta sL^{man}_{ct} &= \beta_0 + \delta_0 t2_t + \beta_1 \Delta IPW^{USCH}_{ct} + \beta_2 (t2_t \times \Delta IPW^{USCH}_{ct}) + \Delta X_{ct} \\ &\quad + \rho_1 \hat{e}^1_{ct} + \rho_2 \hat{e}^2_{ct} + u_{ct}\end{aligned}$$

where $\hat{e}^1_{ct}$ and $\hat{e}^2_{ct}$ are the residuals from the first stage regressions. The null hypothesis of exogeneity is $H_0 : \rho_1 = \rho_2 = 0$. The F statistics is calculated to 8.23. The relevant critical value is $c = 3$ (the 95th percentile of an F distribution with $df = (2, \infty)$). The null hypothesis is therefore rejected and it is concluded that ΔIPW^{USCH} and $t2_t \times \Delta IPW^{USCH}$ are endogenous regressors. As such the IV estimator is the preferred estimator.

Problem 4

1. The regression model is:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i^* + u_i \\ &= \beta_0 + \beta_1 x_i + (u_i - \beta_1 \epsilon_i) \end{aligned}$$

   The asymptotic bias of the IV estimator is (cf. Wooldridge [15.19]):

$$\begin{aligned}
plim\, \widehat{\beta}_1^{IV} - \beta_1 &= \frac{cov(z, u - \beta_1 \epsilon)}{cov(z, x)} \\
&= -\beta_1 \frac{cov(z, \epsilon)}{cov(z, x)} && [cov(z, u) = 0 \text{ by assumption}] \\
&= -\beta_1 \frac{cov(z^* + \eta, \epsilon)}{cov(z^* + \eta, x)} && [\text{Definition of } z] \\
&= -\beta_1 \frac{cov(\eta, \epsilon)}{cov(z^*, x) + cov(\eta, x)} && [cov(z^*, \epsilon) = 0] \\
&= -\beta_1 \frac{cov(\eta, \epsilon)}{cov(z^*, x^*) + cov(z^*, \epsilon) + cov(\eta, x^*) + cov(\eta, \epsilon)} && [\text{Def. of } x] \\
&= -\beta_1 \frac{cov(\eta, \epsilon)}{cov(z^*, x^*) + cov(\eta, \epsilon)} && [\text{Since } cov(z^*, \epsilon) = cov(\eta, x^*) = 0] \\
&= -\beta_1 \frac{\rho \sigma^2_\epsilon}{\theta \sigma^2_{x^*} + \rho \sigma^2_\epsilon}
\end{aligned}$$

The asymptotic bias of the IV estimator depends on the parameters β_1, ρ and θ as well as the variances of the true regressor, x^*, and the measurement error, ϵ. In the special case of uncorrelated measurement errors, i.e. $\rho = 0$, the IV estimator is consistent, $plim\,\widehat{\beta}_1^{IV} = \beta_1$. That is, the IV estimator may remove the attenuation bias as long as the measurement errors in the explanatory variable and the instrument are not correlated.

2. The OLS and IV estimators are equally biased when [$plim\,\widehat{\beta}_1^{OLS}$ defined in W [9.33]]:

$$\begin{aligned} plim\,\widehat{\beta}_1^{IV} &= plim\,\widehat{\beta}_1^{OLS} \\ -\beta_1 \frac{\rho\sigma_\epsilon^2}{\theta\sigma_{x^*}^2 + \rho\sigma_\epsilon^2} &= -\beta_1 \frac{\sigma_\epsilon^2}{\sigma_{x^*}^2 + \sigma_\epsilon^2} \\ \rho(\sigma_{x^*}^2 + \sigma_\epsilon^2) &= \theta\sigma_{x^*}^2 + \rho\sigma_\epsilon^2 \\ \rho\sigma_{x^*}^2 &= \theta\sigma_{x^*}^2 \\ \rho &= \theta \end{aligned}$$

In situations where the instrument is as relevant as it is invalid ($\rho = \theta$), the IV estimator is exactly as biased as the OLS estimator.

Problem 5

The simulation results confirm the theoretical properties derived above:

1. The IV estimator is consistent if $\rho = 0$ as the average $\widehat{\beta}_1^{IV}$ is equal to $\beta_1 = 3$.
2. A negative bias is introduced in the IV estimator when $\rho = 0.5$ (positively correlated measurement errors). Given the stated assumptions, the asymptotic bias is confirmed to be $plim\,\widehat{\beta}_1^{IV} - \beta_1 = -3 \times \frac{0.5\times1}{1\times4+0.5\times1} = -\frac{1}{3} = -0.333$. While the IV estimator is biased in this case, it is less biased compared to the OLS estimator which suffers from a stronger attenuation bias (that is unrelated to ρ).
3. The OLS and IV estimators are equally biased if $\rho = \theta = 1$, cf. the answer to Problem 4. This result extends to cases where $\rho = \theta = c$, where c is a real number.
4. The IV estimator is positively biased if $\rho = -0.5$. This result illustrates that the bias in the IV estimator may be positive or negative depending on the signs of β_1, ρ, θ and the variance terms, $\sigma_{x^*}^2$ and σ_ϵ^2 following the expressions in Problem 4. Measurement errors may, in other words, lead to positive bias in the IV estimator, i.e., bias away from zero. This contrasts with the attenuation bias observed for the OLS estimator which always lead to bias towards zero (no matter the sign of β_1).

Table 7: MC results

		Mean	Std.dev.	Min	Max
$\rho = \frac{1}{2}$:	OLS	2.400	0.039	2.291	2.541
$\rho = \frac{1}{2}$:	IV	2.667	0.046	2.537	2.858
$\rho = 0$:	IV	3.001	0.054	2.832	3.189
$\rho = 1$:	IV	2.399	0.043	2.279	2.580
$\rho = -\frac{1}{2}$:	IV	3.431	0.074	3.199	3.710