Note for *Market Access and Individual Wages: Evidence from China*
Laura Hering and Sandra Poncet

Data Sources

1. Trade data used for market access calculation
International trade flows are obtained from the IMF Direction of Trade Statistics (DOTS) and can be bought on the IMF website (http://www.imf.org/external/data.htm#data).

Internal trade flows are calculated as the difference between domestic primary and secondary sector production minus exports. Production data for OECD countries come from the OECD STAN database (available for a fee). For other countries, ratios of industry and agriculture output in percentages of GDP are extracted from Datastream (http://www.lib.uwo.ca/business/data.html). They are then multiplied by the countries' GDP from World Development Indicators 2005 (can be purchased on the website of the World Bank).

Intra-provincial trade flows for China and intra-national trade flows for foreign countries are computed, as domestic production minus total exports. Total production for Chinese provinces is computed as the sum of industrial and agricultural output.

Provincial foreign trade data are obtained from the Customs General Administration database and can be purchased at http://www.allmyinfo.com/eng/Services/index1-3.asp

2. Province and city-level data
Information on provincial and city characteristics (gdp, area, population) stem from the China Statistical Yearbook and can be downloaded for a fee at the China data center of the University of Michigan (http://chinadatacenter.org/newcdc/).

3. Individual data
Individual and household data come from the 1995 survey of the China Household Income Project (CHIP), which is a joint research effort sponsored by the Institute of Economics, at the Chinese Academy of Social Sciences, the Asian Development Bank and the Ford Foundation. Additional support was provided by the East Asian Institute, Columbia University. More detailed description is available at
http://www.icpsr.umich.edu/cocoon/ICPSR/STUDY/03012.xml.
At the same address, the data set can be downloaded for a fee.

Data sets

All data sets are in Stata 9.0 format (Windows).

The data set *indvidualdata.dta* contains the 6,848 individuals used in our analysis (footnote 25 of the paper explains how we obtain this number). Variables include all required individual and household characteristics as well as city data and our calculated indicators at the city level.

The data set *final.dta* is the final data set which contains all necessary information. .do files *table1-table5* and *AdditionalResults* should be run on this file.

Since trade data has to be purchased, we cannot publish our trade data used for the estimation of the gravity equation.

.do files

The .do file *individualdata.do* uses *indvidualdata.dta*. It calculates our dependent variables (the log of wages per hour), the interaction variables and the additional variables for the selection equation of the Heckman estimations. Running this file permits to reconstruct the data set data set *final.dta.*

The .do files *table1.do, table2.do, table3.do, table4.do* and *table5.do* replicate the results of the five tables of the section "Tables".

AdditionalResults.do gives the additional results (our benchmark regression with different wage definitions and the heckman regressions testing for a selection bias in our sample) which are available on the corresponding author's website (http://ces.univ-paris1.fr/membre/Poncet/).

In the paper we claim to perform the bootstrap procedure to all our regressions. Since standard errors of our dependent variables didn't change significantly, tables reported and replicated in the available .do files rely on OLS estimations.

Contact details of the authors

Sandra Poncet: sandra.poncet@univ-paris1.fr
Laura Hering: laura.hering@gmail.com

Data Dictionnary (file final.dta)

Variables at the individual level (source: CHIP 1995)

n1	Identifiant of household
a2	Identifiant of individual in the household
a3	Relationship to Household head
sex	2=woman, 1=man
age	Age
martialstatus	Martial status
Communist	2=no member, 1=member of communist party
education	Educational level
schoolyears	Years of schooling
experience	Years work experience
Sector	Sector of activity
ownership1	Firm ownership for 1st job
ownership2	Firm ownership for 2nd job
occupation	Occupation/ Position within the firm
Hoursday	Work hours on an average day
Daysweek	Average number of work days per week
daysunempl	Days unemployed during the year
totalincome	Total ind. income in 1995
wages	Wage
allowance	Allowances while waiting for job
bonus	Bonus wages
a55	Other allowances and subsidies
a62	Other income from the work unit
a63	Hardship allowances
a71	Property Income
a76	Transfer income
a85	Income from household sideline production
a89	Net income in kind
ville	Name of city
codecity	Code of city
Province	Name of province
county	Code of county
basewage	Annual earnings
basewageh	Wage per hour
lnbasewage	ln(Annual earnings)
lnbasewageh	Dep var: ln(wage per hour)
wage_alt1	Basic salary - other allowances
lnwageh_alt1	ln(Basic salary - other allowances (per hour))
wage_alt2	Basic salary +bonus+sub+ other income
lnwageh_alt2	ln(Basic salary+bonus+sub+other incomes (per hour))
wagehour	Basic salary+bonus+sub+allowances(job) (per hour)
lnwageh_alt3	ln(Basic salary+bonuses+subsidies+allowances(job)

	(per hour))
age2	age*age
dnq	1= unqualified worker (schooly <9), 0 otherwise
communist	1=member of communist party, 0 otherwise
ownstate	State-owned enterprise (central or provincial level)
ownlocal	Local publicly owned enterprise
ownucoll	Urban collective enterprise
ownprivate	Private enterprise
ownselfempl	Self-employed (single person enterprise)
ownsinoforegin	Sino-foreign joint ventures
ownforeign	Foreign-invested enterprises
ownallothers	Other ownership
occupowner	Owner of private or individual enterprise
occupownmanager	Owner and manager of private enterprise
occupprofessional	Professional or technical worker
occupheadInst	Head of institution
occupdivision	Division head in institution
occupoffice	Office worker
occupskilled	Skilled worker
occupunskilled	Unskilled worker
occupallother	Other occupation
outliers	Hadi outlier (p=.01), 1 = outlier, 0 otherwise
female	1=woman, 0 otherwise
male	1=man, 0 otherwise
married	1=married, 0 otherwise
married_male	married*male
married_female	married*female
noworkincome	Income from other sources than work
Variables at the household level (CHIP 1995)	
h1	HH: Total value of all financial assets
h9	HH: Durable consumer goods
h10	HH: Market value of productive fixed assets
h11	HH: Market value of self-owned house
h12	HH: Market value of other assets
h13	HH: debts of household
hhmembers	HH: Household size
hhwealth	HH: Household wealth
Variables at the city level	
pop1997	City: population in 1997
lnpop1997	City: ln(city population in 1997)
lngdp1997	City: GDP in 1997
lnland1997	City: area in 1997
living_costs	Living costs index of the city
skill_int	ln(skill intensity of the city)
spat_dec	ln(Human Cap Spatial decay of city)

lnPopMA	ln(MA weighted by population)
lnLcostsMA	ln(MA weighted by living costs)
lnMA	ln(Market access of the city)
lnharris	ln(Harris (1954) market potential)
lnsuminvRDM1	ln(Instrument1: Centrality Rest of the world)
lnsuminvchin1	ln(Instrument1: Centrality China)
code_o	String var: province & city code
Codeprov	Code of province
Interaction variables	
codeprovsect	Province-sector code
dprovsect1 - dprovsect33	Dummies for province-sector pair
MAPrivate	lnMA*ownprivate
MAState	lnMA*ownstate
MAColl	lnMA*ownucoll
MAForeign	lnMA*ownforeign
MALocal	lnMA*ownlocal
MASinoforeign	lnMA*ownsinoforegin
MAAllothers	lnMA*ownallothers