Testing for Pollution Haven Behavior—Evidence from China

Yang Fan

Data Description:

This dataset includes province-level data for 27 provinces, with the exception of Beijing, Tianjin, Shanghai and Tibet which had missing data from 1996 to 2003. Variables such as road length, railroad length, telephones, and consumption are collected from Chinese Statistical Yearbooks. Data for road, railroad, and telephones of Sichuan province for 1996 are missing. The dependent variable, number of new FDI projects in each province for each year from 1996 to 2003, is collected from provincial statistical yearbooks. The independent variable, average levy, is calculated as total collected water pollution levies over wastewater. Data of total collected water pollution levies and wastewater is collected from the Chinese Environmental Yearbook. Wages by skill level are not available, so it is assumed that relative labor supplies determine relative wages in each province. The number is collected from each year's China Statistical Yearbook and 2000 Population Census, with an overall 1% sample of the population surveyed. Skilled labor is defined as those with senior secondary or higher education, semi-skilled labor as those with primary and junior secondary level education, and unskilled labor as those are illiterate or have less than primary level education. FDI incentives are constructed as a dummy, with provinces that have a special economic zone (SEZ) or open coastal city (OCC) in the province labeled one. All right hand variables except FDI incentives and ratio of skilled and unskilled labor, are lagged one year to represent predetermined information, available to an investor at the time of the location decision. Information of FDI incentive and labor education level can be observed at the time of the location decision, and are changing more frequently relative to other infrastructure measures. In addition, these two controls are uncorrelated with environmental regulation enforcement. The following variables are included:

variable name	storage type	display format	value label	variable label
id	byte	%8.0g		Unique Provincial ID
province	str14	%14s		Province
year	int	%8.0g		Year (1996-2003)
region	str9	%9s		Region
r_id	byte	%8.0g		Unique Regional ID
fdi	int	%8.0g		FDI New Projects
lag_tcwpl	float	%9.0g		Total collected water pollution levies (10000 RMB, lagged one year)
lag_ww	float	%9.0g		Industrial Waste Water Discharged (100 million tons,lagged one year)
lag_railroad	float	%9.0g		Railroads (10000km,lagged one year)
lag_roads	str5	%9s		Roads (10000km, lagged one year)
lag_tele	float	%9.0g		Telephones (10000 units,lagged one year)
lag_pop	float	%9.0g		Population (10000, lagged one year)
lag_con	float	%9.0g		Consumption(100 million yuan, lagged one year)
sezorocc	byte	%8.0g		SEZ or OCC
pidrity	byte	%8.0g		Provincial Initial Environmental Condition (dirty), =1 if initially above avg
area	float	%9.0g		Area (sq.km)
lag_aclofw	float	%9.0g		Average Levies of Wastewater (yuan/ton, lagged one year)
illterate	long	%12.0g		Illterate
primary	long	%12.0g		Primary
juniorss	long	%12.0g		Junior SS
seniorss	long	%12.0g		Senior SS
collegeorhigher	long	%12.0g		College or higher