Four New Empirical Tests of the Pollution Haven Hypothesis When Environmental Regulation is Endogenous

Daniel L. Millimet[1] Jayjit Roy[2]

[1]Southern Methodist University & IZA

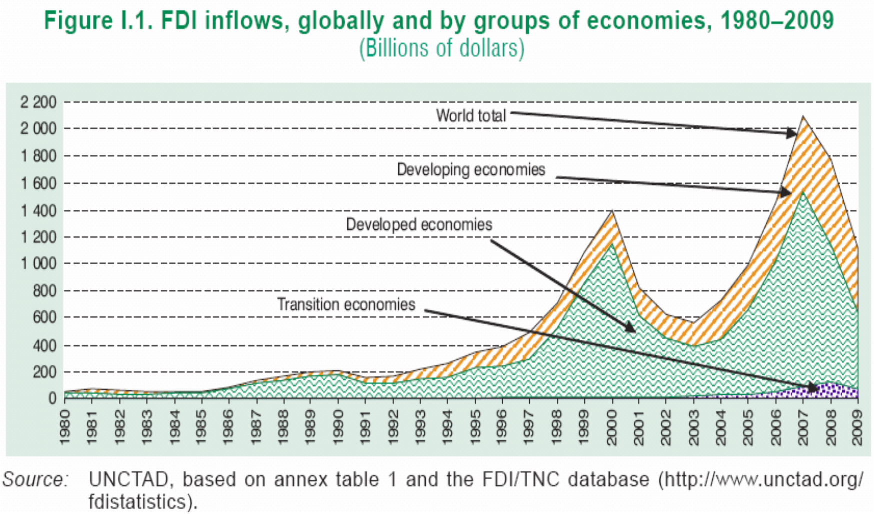
[2]Appalachian State University

March 2011

Introduction

- Pollution Haven Hypothesis (PHH): reduced trade barriers enable polluting multinational enterprises (MNEs) to outsource production activities to areas with lax environmental regulation
- Why do we care?
  - Governments may attempt to attract foreign direct investment (FDI) at the cost of environment $\xrightarrow{?}$ scope for international coordination
  - Countries may use the degree of environmental stringency to influence trade patterns $\xrightarrow{?}$ scope for incorporating environmental rules into trade agreements
  - Existing trade agreements (e.g., WTO) may prevent governments from choosing a desired regulatory regime due to 'unintended' effect on trade patterns
- Hence, an empirical examination of the validity of the PHH is extremely relevant from a policy perspective

- Understanding FDI determinants is also important in its own right; tremendous growth of late



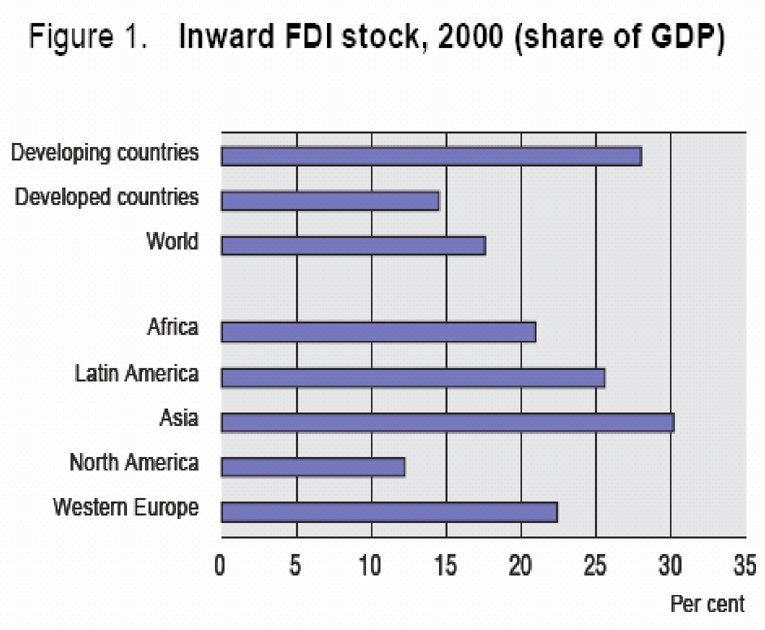
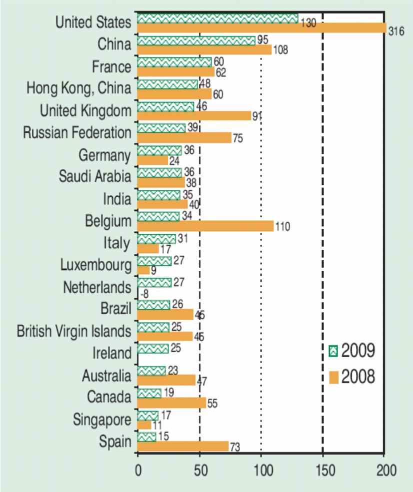


Figure 1. Inward FDI stock, 2000 (share of GDP)
Developing countries
Developed countries
World
Africa
Latin America
Asia
North America
Western Europe
0
5
10
15
20
25
30
35
Per cent

- U.S. is largest host of global FDI

Figure I.4. Global FDI inflows, top 20 host economies, 2008–2009[a]
(Billions of dollars)



Source: UNCTAD, based on annex table 1 and the FDI/TNC database (http://www.unctad.org/fdistatistics).

Literature Review

- Environmental stringency *does* affect FDI location decisions
  - Kellenberg (2009), Fredriksson et al. (2003), Keller and Levinson (2002), List and Co (2000)
- Environmental stringency *does not* affect FDI location decisions
  - List et al. (2004), Eskeland and Harrison (2003)
- Heterogeneous effects of environmental stringency on FDI location decisions
  - Dean et al. (2009) ... country of origin, pollution intensity
  - Henderson and Millimet (2006) ... nonparametric estimates

- Potential shortcomings of the existing literature
  1. Environmental regulation is difficult to *measure*
     - ★ Measurement error bias can be exacerbated in fixed effects model (common in literature)
  2. Typically, *spatial effects* are not controlled for or done so in a theoretically inconsistent way
     - ★ Third-country effects recently shown to be quite relevant to FDI location decisions (Blonigen et al. 2007; Baltagi et al. 2007, 2008)
     - ★ Conditional logit models are too restrictive
  3. Very few attempts to address the issue of *endogeneity* of environmental stringency
     - ★ Kellenberg 2009; Levinson and Taylor 2008; Ederington and Minier 2003; Fredriksson et al. 2003
     - ★ Even so, instruments are suspect

Contribution

- To fix ideas, typical model of PHH

$$\ln(FDI_{it}) = \eta_i + \lambda_t + x_{it}\beta + \tau R_{it} + \varepsilon_{it}$$

where FDI_{it} is inbound FDI to location i in time t, η and λ are location and time effects, x are controls, and R is a proxy for environmental regulation

- Our contribution:
  1. Incorporate spatial effects into an empirical model of the PHH
  2. Address endogeneity and measurement error using four novel identification strategies
     - ★ Differencing IV strategy based on Pitt and Rosenzweig (1990)
     - ★ Higher Moments IV strategy of Lewbel (2010)
     - ★ Control Function approach of Klein and Vella (2009)
     - ★ Covariance Restrictions strategy of Chamberlain and Griliches (1975)

Data

- From Keller and Levinson (2002)
- Cover the 48 contiguous U.S. states from 1977-1994, omitting 1987
- Two FDI measures
  - Value of gross property, plant, and equipment of foreign-owned affiliates
  - Employment at foreign-owned affiliates
- Sectors
  - All manufacturing
  - Chemicals
  - Metals
  - Food and Kindred Products
  - Machinery
- Measure of regulation: Levinson's (2001) measure of relative abatement costs (RAC)
- Covariates: average production-worker wages, land prices, energy prices, total road mileage, unemployment rate, unionization rate, market proximity, population, tax effort

- Levinson's (2001) RAC measure

$$R_{it} = \frac{PACE_{it} / Y_{it}}{(\widehat{PACE_{it} / Y_{it}})}$$

where

  - $PACE$ = pollution abatement costs
  - Y = gross state product (GSP)
  - $(\widehat{PACE_{it} / Y_{it}})$ = predicted PACE per unit of GSP; i.e.,

$$(\widehat{PACE_{it} / Y_{it}}) = \frac{1}{Y_{it}} \sum_{m=20}^{39} \left(\frac{Y_{imt}}{Y_{it}} \right) PACE_{mt}$$

where $m = 20, ..., 39$ indexes 2-digit SIC manufacturing sectors

- Hence, R is the ratio of *actual* abatement costs per dollar of output to *predicted* abatement costs per dollar of output given the industrial composition of states
- State is

$$R_{it} = \begin{cases} > 1 & \text{stringent} \\ = 1 & \text{average} \\ < 1 & \text{lax} \end{cases}$$

State	Abatement Cost Index S^* (1)	Unadjusted Index S (2)
Alabama	1.19	0.0219
Arizona	1.39	0.0148
Arkansas	1.17	0.0168
California	0.90	0.0121
Colorado	1.01	0.0113
Connecticut	0.67	0.0079
Delaware	1.30	0.0344
Florida	1.21	0.0138
Georgia	0.91	0.0127
Idaho	1.66	0.0181
Illinois	0.91	0.0132
Indiana	1.14	0.0196
Iowa	0.96	0.0106
Kansas	0.76	0.0115
Kentucky	0.99	0.0146
Louisiana	1.51	0.0538
Maine	1.55	0.0237
Maryland	1.17	0.0185
Massachusetts	0.67	0.0067
Michigan	1.01	0.0121
Minnesota	0.66	0.0092
Mississippi	1.47	0.0213
Missouri	0.79	0.0104
Montana	1.49	0.0341
Nebraska	0.83	0.0088
Nevada	0.63	0.0072
New Hampshire	0.75	0.0072
New Jersey	0.82	0.0158
New Mexico	1.64	0.0306
New York	0.77	0.0087
N. Carolina	0.82	0.0088
N. Dakota	0.77	0.0105
Ohio	0.82	0.0139
Oklahoma	0.58	0.0103
Oregon	1.22	0.0139
Pennsylvania	0.91	0.0169
Rhode Island	0.72	0.0075
S. Carolina	0.99	0.0160
S. Dakota	0.68	0.0056
Tennessee	1.10	0.0165
Texas	1.39	0.0311
Utah	0.93	0.0164
Vermont	0.66	0.0065
Virginia	0.96	0.0118
Washington	1.37	0.0196
W. Virginia	1.58	0.0433
Wisconsin	0.89	0.0110
Wyoming	0.72	0.0259
Avg. for lowest 5[b]	0.64	0.0082
Avg. for highest 5[c]	1.59	0.0339
Avg. for lowest 20[d]	0.75	0.0103
Avg. for highest 20[e]	1.33	0.0235

Empirical Model

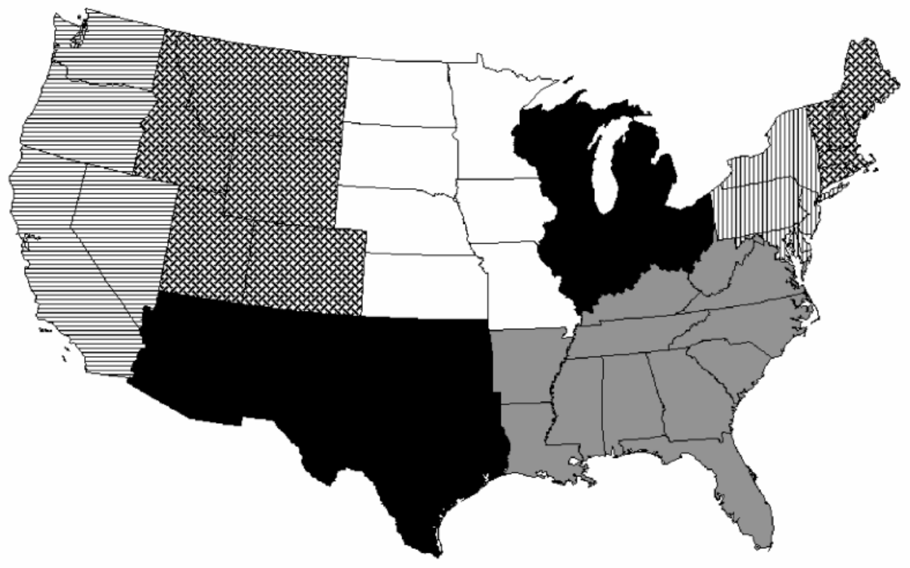
- 'Structural' model is

$$\begin{aligned}\ln(FDI_{it}) &= \eta_i + \lambda_t + \\ &+ \sum_k \left[\underbrace{\beta_k \ln(x_{ikt})}_{\text{own effects}} + \underbrace{\delta_k \ln\left(\sum_j \omega_{ij} x_{jkt}\right)}_{\text{spatial effects}} \right] + \varepsilon_{it}\end{aligned}$$

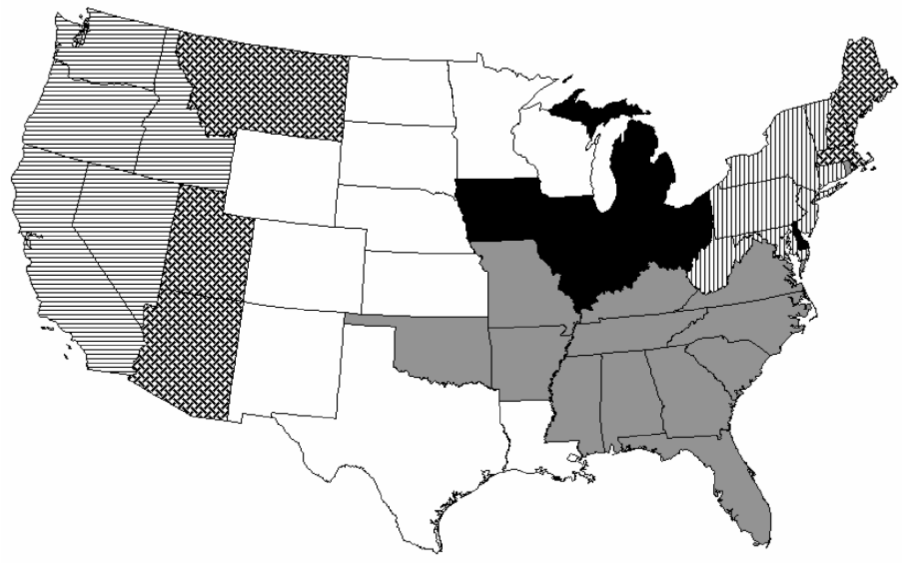
  where k indexes covariates and R is one such covariate, and ω_{ij} weight attached to state j by state i
- FE estimation requires all x's be strictly exogenous
- Not likely for R_{it}, $\sum_j \omega_{ij} R_{jt}$
  - ε_{it} captures local political activism, corruption, transboundary pollution, measurement error, ...
  - R_{it} may respond to past shocks impacting FDI
- Need an alternative identification strategy

- Before turning to identification, a few comments on the weights, ω
  - The weights must be specified
  - Specification of the weights is admittedly ad hoc, but
    - ★ In general, this is really just an issue of proper interpretation: estimates are a lower bound of spatial effects
    - ★ In other words, mis-specification of the weights should be thought of as measurement error
    - ★ Solutions: sensitivity to different weighting schemes, or IV
- Here, we use three weighting schemes
  1. Contiguous neighbors
  2. BEA regional neighbors
  3. Crone (1998/1999) regional definition

BEA Regions



Crone Regions



Identification Strategy #1

Pitt & Rosenzweig (1990) - Type Approach

- Original setup
  - Assess differential effect of endogenous household-level variable (infant illness) on children (health and time allocation) by gender
  - So, setup entails two types of children and a common aggregate (household) variable with possible differential returns
  - Absent typical instruments, solution proposed is to estimate the *differential* effect of the endogenous variable on boys versus girls, generating instruments by assuming some exogenous variables have equal effects on boys and girls
- Our setup
  - Assess *differential* effect of endogenous state-level variable (regulation) on two types of FDI
  - Here, we partition sectors into pollution-intensive (PI) and non-intensive (NPI) sectors
  - Instruments are generated by assuming some exogenous variables have equal effects on PI and NPI sectors

- US inbound FDI in PI manufacturing sectors

$$
\begin{aligned}
\ln(FDI_{it}^{p}) &= \widetilde{\eta}_i^{p} + \widetilde{\lambda}_t^{p} \\
&+ \sum_{k=1}^{K_1} \left[\beta_{1k}^{p} \ln(x_{1ikt}) + \delta_{1k}^{p} \ln\left(\textstyle\sum_j \omega_{ij} x_{1jkt}\right)\right] \\
&+ \sum_{k=1}^{K_2} \left[\beta_{2k} \ln(x_{2ikt}) + \delta_{2k} \ln\left(\textstyle\sum_j \omega_{ij} x_{2jkt}\right)\right] + \varepsilon_{it}^{p}
\end{aligned}
$$

where

- FDI_{it} : aggregate inbound FDI into state i at time t; $\widetilde{\eta}_i$ are state effects; $\widetilde{\lambda}_t$ are time effects
- x_{1ikt} : average production-worker wages, land prices, energy prices, total road mileage, unemployment rate, unionization rate, RAC in state i at time t; RAC denotes relative abatement cost
- x_{2ikt} : market proximity, population, tax effort in state i at time t

- US inbound FDI in NPI manufacturing sectors

$$
\begin{aligned}
\ln(FDI_{it}^{np}) &= \widetilde{\eta}_i^{np} + \widetilde{\lambda}_t^{np} \\
&+ \sum_{k=1}^{K_1} \left[\beta_{1k}^{np} \ln(x_{1ikt}) + \delta_{1k}^{np} \ln\left(\textstyle\sum_j \omega_{ij} x_{1jkt}\right)\right] \\
&+ \sum_{k=1}^{K_2} \left[\beta_{2k} \ln(x_{2ikt}) + \delta_{2k} \ln\left(\textstyle\sum_j \omega_{ij} x_{2jkt}\right)\right] + \varepsilon_{it}^{np}
\end{aligned}
$$

where

- FDI_{it} : aggregate inbound FDI into state i at time t; $\widetilde{\eta}_i$ are state effects; $\widetilde{\lambda}_t$ are time effects
- x_{1ikt} : average production-worker wages, land prices, energy prices, total road mileage, unemployment rate, unionization rate, RAC in state i at time t; RAC denotes relative abatement cost
- x_{2ikt} : market proximity, population, tax effort in state i at time t

- Subtracting the NPI equation from PI equation ($\Delta \Rightarrow$ PI$-$NPI)

$$\begin{aligned} \Delta \ln(FDI_{it}) &= \eta_i + \lambda_t + \\ &+ \sum_{k=1}^{K_1} \left[\Delta\beta_{1k} \ln(x_{1ikt}) + \Delta\delta_{1k} \ln\left(\textstyle\sum_j \omega_{ij} x_{1jkt}\right)\right] + \Delta\varepsilon_{it} \end{aligned}$$

- Notes:
  - $\ln(x_{2ikt})$ and $\ln\left(\sum_j \omega_{ij} x_{2jkt}\right)$, $k = 1, ..., K_2$, are available as exclusion restrictions
  - Identification requires choosing x_2 and imposing $\beta_2^p = \beta_2^{np} = \beta_2$
  - $K_2 > 1 \Rightarrow$ the usual overidentification test constitutes a test of the restrictions imposed
  - η_i are replaced with region effects ($\widetilde{\eta}_i^p - \widetilde{\eta}_i^{np}$ is assumed to be constant within regions)
  - Unless environmental stringency is assumed to have no effect on FDI in the NPI sector, only *differential* effects of own and neighboring environmental stringency on PI FDI *relative* to NPI FDI are identified

Results

Pitt & Rosenzweig (1990) - Type Approach

Determinants of Relative FDI: Region Effects & Time Effects

	PP&E							
	Spec (1)		Spec (2)		Spec (3)		Spec (4)	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
ln(index)	-0.789*	-4.720*	-0.724*	-4.657*	-0.802*	-3.437*	-0.657*	-3.535*
	(0.155)	(0.653)	(0.146)	(0.703)	(0.142)	(0.509)	(0.151)	(0.631)
ln(spatial index)			-1.056*	0.930	0.596	0.977	0.148	7.210†
			(0.349)	(01.718)	(0.498)	(01.958)	(0.576)	(3.636)
Underid Test		0.000		0.000		0.000		0.000
F-stat		39.174		8.674		6.344		9.340
Overid Test		0.040		0.136		0.074		0.032
Endogeity		0.000		0.000		0.000		0.000
Joint Sign. Endog.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
N	563	563	563	563	563	563	563	563

‡ $p<0.10$, † $p<0.05$, * $p<0.01$. Underid reports the p-value of the Kleibergen-Paap (2006) rk statistic with rejection implying identification; Overid reports the p-value of Hansen J statistic with rejection casting doubt on instruments' validity; Endog reports the p-value of endogeneity test of the endogenous regressors; Jsig reports the p-value of Anderson-Rubin chi-square test of endogenous regressors; F-stat reports the Kleibergen-Paap F statistic for weak identification.

Determinants of Relative FDI: Region Effects & Time Effects

	Employment							
	Spec (1)		Spec (2)		Spec (3)		Spec (4)	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
ln(index)	-0.498*	-2.205*	-0.447*	-2.399*	-0.494*	-1.463*	-0.584*	-2.212*
	(0.096)	(0.304)	(0.096)	(0.339)	(0.099)	(0.308)	(0.101)	(0.346)
ln(spatial index)			-0.758*	1.052	0.096	1.984†	-0.546	2.421
			(0.211)	(0.869)	(0.277)	(0.929)	(0.398)	(1.903)
Underid		0.000		0.000		0.000		0.000
F-stat		38.003		13.256		10.666		10.037
Overid		0.834		0.210		0.172		0.164
Endog		0.000		0.000		0.000		0.000
JSig	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
N	621	621	621	621	621	621	621	621

‡ p<0.10, † p<0.05, * p<0.01. Underid reports the p-value of the Kleibergen-Paap (2006) rk statistic with rejection implying identification; Overid reports the p-value of Hansen J statistic with rejection casting doubt on instruments' validity; Endog reports the p-value of endogeneity test of the endogenous regressors; Jsig reports the p-value of Anderson-Rubin chi-square test of endogenous regressors; F-stat reports the Kleibergen-Paap F statistic for weak identification.

Results

Pitt & Rosenzweig (1990) - Type Falsification Test

- Additional check on the identification strategy
- Intuition
  - If the strategy is valid and correctly identifies a significant effect of environmental stringency in PI sectors relative to NPI sectors, then the corresponding coefficients should be insignificant when two NPI sectors are compared
  - Also, if environmental stringency is correctly identified to have no significant effects in PI sectors relative to NPI sectors, then the corresponding coefficients should continue to remain insignificant when two NPI sectors are compared
- 2 NPI sectors considered: Food and Kindred Products and all other *non-chemical* manufacturing sectors
- Effects of own and neighboring stringency levels are mostly insignificant in IV models where instruments pass relevant specification tests

Identification Strategy #2

Lewbel (2010) - Type Approach

- Lewbel (2010) considers identification in triangular and simultaneous equation systems
- Typical solution is IV
- Absent typical instruments, solution utilizes higher moments of the data
- Similar estimators proposed in Lewbel (1997), Ebbes et al. (2004)

- Model

$$\ln(FDI_{it}) = \eta_i + \lambda_t + \sum_{k=1}^{K} \left[\beta_k \ln(x_{ikt}) + \delta_k \ln\left(\sum_j \omega_{ij} x_{jkt}\right)\right] + \varepsilon_{it}$$

- IVs utilized for R_{it} and $\sum_j \omega_{ij} R_{jt}$ are

$$\widetilde{z}_{ijt} = (z_{it} - \overline{z})\widehat{\zeta}_{ijt}, \quad j = 1, 2$$

  where

  - ζ_j is the error in the first-stage regression for endogenous regressor j
  - $z \subseteq x$ satisfying

$$\begin{aligned} \mathbb{E}[z'\zeta_j^2] &\neq 0 \quad \text{(correlated with error variance)} \\ \mathbb{E}[z'\varepsilon\zeta_j] &= 0 \quad \text{(uncorrelated with error covariance)} \end{aligned}$$

    $j = 1, 2$, which is feasible if endogeneity is due to a common unobserved factor
  - In practice, z includes land prices, total road mileage, and market proximity (and spatial analogs)

Results

Lewbel (2010) - Type Approach

Determinants of FDI: State Effects & Time Effects

	Chemical Sector (PP&E)							
	Spec (1)		Spec (2)		Spec (3)		Spec (4)	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
ln(index)	-0.198†	-0.558*	-0.199†	-0.218	-0.153	-0.419†	-0.222†	-0.384‡
	(0.092)	(0.195)	(0.092)	(0.168)	(0.095)	(0.177)	(0.098)	(0.209)
ln(spatial index)			-0.264	-0.205	-0.313	-0.978†	0.323	0.683
			(0.169)	(0.445)	(0.208)	(0.409)	(0.310)	(0.837)
Underid Test		0.000		0.000		0.000		0.000
F-stat		41.951		7.885		11.003		6.367
Overid Test		0.814		0.001		0.027		0.166
Endogeity		0.036		0.984		0.175		0.431
Joint Sign. Endog.	0.033	0.03	0.023	0.001	0.128	0.004	0.018	0.04
N	563	563	563	563	563	563	563	563

‡ p<0.10, † p<0.05, * p<0.01. Underid reports the p-value of the Kleibergen-Paap (2006) rk statistic with rejection implying identification; Overid reports the p-value of Hansen J statistic with rejection casting doubt on instruments' validity; Endog reports the p-value of endogeneity test of the endogenous regressors; Jsig reports the p-value of Anderson-Rubin chi-square test of endogenous regressors; F-stat reports the Kleibergen-Paap F statistic for weak identification.

Determinants of FDI: State Effects & Time Effects

	Chemical Sector (Employment)							
	Spec (1)		Spec (2)		Spec (3)		Spec (4)	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
ln(index)	-0.397*	-0.802*	-0.386*	-0.524*	-0.291*	-0.563*	-0.379*	-0.896*
	(0.072)	(0.135)	(0.072)	(0.131)	(0.071)	(0.124)	(0.077)	(0.168)
ln(spatial index)			-0.343†	-0.143	-0.273‡	-0.771*	0.270	-0.733
			(0.134)	(0.375)	(0.153)	(0.288)	(0.254)	(0.771)
Underid Test		0.000		0.000		0.000		0.000
F-stat		68.248		10.028		13.998		5.675
Overid Test		0.137		0.000		0.030		0.392
Endogeity		0.000		0.362		0.039		0.001
Joint Sign. Endog.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
N	621	621	621	621	621	621	621	621

‡ p<0.10, † p<0.05, * p<0.01. Underid reports the p-value of the Kleibergen-Paap (2006) rk statistic with rejection implying identification; Overid reports the p-value of Hansen J statistic with rejection casting doubt on instruments' validity; Endog reports the p-value of endogeneity test of the endogenous regressors; Jsig reports the p-value of Anderson-Rubin chi-square test of endogenous regressors; F-stat reports the Kleibergen-Paap F statistic for weak identification.

Determinants of FDI: State Effects & Time Effects

	Total Manufacturing (PP&E)							
	Spec (1)		Spec (2)		Spec (3)		Spec (4)	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
ln(index)	-0.079‡	-0.155†	-0.082‡	-0.241*	-0.011	-0.043	-0.086‡	-0.109
	(0.046)	(0.073)	(0.047)	(0.083)	(0.046)	(0.075)	(0.048)	(0.080)
ln(spatial index)			-0.126	-0.944*	-0.094	-0.354‡	0.181	0.387
			(0.096)	(0.300)	(0.100)	(0.193)	(0.159)	(0.446)
Underid Test		0.000		0.000		0.000		0.000
F-stat		137.812		10.218		20.012		8.860
Overid Test		0.758		0.000		0.001		0.211
Endogeity		0.195		0.043		0.345		0.702
Joint Sign. Endog.	0.086	0.168	0.089	0.000	0.638	0.000	0.057	0.110
N	811	811	811	811	811	811	811	811

‡ p<0.10, † p<0.05, * p<0.01. Underid reports the p-value of the Kleibergen-Paap (2006) rk statistic with rejection implying identification; Overid reports the p-value of Hansen J statistic with rejection casting doubt on instruments' validity; Endog reports the p-value of endogeneity test of the endogenous regressors; Jsig reports the p-value of Anderson-Rubin chi-square test of endogenous regressors; F-stat reports the Kleibergen-Paap F statistic for weak identification.

Determinants of FDI: State Effects & Time Effects

	Total Manufacturing (Employment)							
	Spec (1)		Spec (2)		Spec (3)		Spec (4)	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
ln(index)	-0.013	0.042	-0.027	0.054	0.053	0.192*	-0.042	-0.024
	(0.041)	(0.065)	(0.041)	(0.066)	(0.040)	(0.063)	(0.043)	(0.071)
ln(spatial index)			0.039	0.131	-0.023	-0.041	0.201	-0.421
			(0.086)	(0.237)	(0.089)	(0.162)	(0.144)	(0.407)
Underid Test		0.000		0.000		0.000		0.000
F-stat		147.699		11.265		22.029		8.927
Overid Test		0.714		0.000		0.031		0.219
Endogeity		0.283		0.352		0.009		0.109
Joint Sign. Endog.	0.754	0.778	0.731	0.000	0.396	0.002	0.158	0.273
N	814	814	814	814	814	814	814	814

‡ p<0.10, † p<0.05, * p<0.01. Underid reports the p-value of the Kleibergen-Paap (2006) rk statistic with rejection implying identification; Overid reports the p-value of Hansen J statistic with rejection casting doubt on instruments' validity; Endog reports the p-value of endogeneity test of the endogenous regressors; Jsig reports the p-value of Anderson-Rubin chi-square test of endogenous regressors; F-stat reports the Kleibergen-Paap F statistic for weak identification.

Identification Strategy #3

Klein & Vella (2009) Approach

- Structural Model

$$\begin{aligned} \ln(FDI_{it}) &= X_{it}\Pi + \tau_1 R_{it} + \tau_2 \overset{\vee}{R}_{it} + \varepsilon_{it} \\ R_{it} &= X_{it}\Pi_1 + \zeta_{1it} \\ \overset{\vee}{R}_{it} &= X_{it}\Pi_2 + \zeta_{2it} \end{aligned}$$

  where R and $\overset{\vee}{R}$ are own and spatially lagged RAC, X includes all other regressors (own and spatially lagged exogenous vars and state and time FEs)

- Identification relies on the assumption of heteroskedastic errors, but constant correlation
- As in Lewbel (2010) approach, this is feasible if endogeneity is due to a common unobserved (homoskedastic) factor

- Assumptions
  - $\mathbb{E}[\varepsilon_{it}|X_i] = \mathbb{E}[\zeta_{jit}|X_i] = 0$, $j = 1, 2$
  - $\varepsilon_{it} = S_\varepsilon(X_{it})\,\varepsilon^*_{it}$
  - $\zeta_{jit} = S_j(X_{it})\zeta^*_{jit}$, $j = 1, 2$
  - $S_\varepsilon(X_{it})\,/\,S_j(X_{it})$, $j = 1, 2$, are not contant across i (crucial)
  - $\mathbb{E}(\varepsilon^*_{it}\zeta^*_{jit}) = \rho_j$, $j = 1, 2$

  where ε^*_{jit}, ζ^*_{1it}, and ζ^*_{2it} are homoskedastic errors
- Heteroskedasticity specification ($z_j \subseteq X$, $j = \varepsilon, 1, 2$)

$$\begin{aligned} S^2_{\varepsilon it} &= \exp\left(z_{\varepsilon it}\theta_\varepsilon\right) \\ S^2_{jit} &= \exp\left(z_{jit}\theta_j\right), j = 1, 2 \end{aligned}$$

- In practice
  - z_j includes land prices, total road mileage, and market proximity (as in Lewbel-type approach)
  - z_ε includes market proximity, wages, and population (chemicals) or market proximity (total manufacturing)

- Define

$$\begin{aligned} \widetilde{\varepsilon}_{it} &\equiv \varepsilon_{it} - A(X_{it})\,\zeta_{1it} - B(X_{it})\,\zeta_{2it} \\ A(X_{it}) &\equiv \frac{\mathrm{Cov}(\varepsilon_{it}, \zeta_{1it}|X_i)}{\mathrm{Var}(\zeta_{1it}|X_i)} = \rho_1 \frac{S_\varepsilon(X_{it})}{S_1(X_{it})} \\ B(X_{it}) &\equiv \frac{\mathrm{Cov}(\varepsilon_{it}, \zeta_{2it}|X_i)}{\mathrm{Var}(\zeta_{2it}|X_i)} = \rho_2 \frac{S_\varepsilon(X_{it})}{S_2(X_{it})} \end{aligned}$$

- The FDI equation can now be re-written as

$$\ln(FDI_{it}) = X_{it}\Pi + \tau_1 R_{it} + \tau_2 \overset{\vee}{R}_{it} + \underbrace{\rho_1 \frac{S_\varepsilon(X_{it})}{S_1(X_{it})}\zeta_{1it} + \rho_2 \frac{S_\varepsilon(X_{it})}{S_2(X_{it})}\zeta_{2it}}_{\text{control function}} + \widetilde{\varepsilon}_{it}$$

- Estimation
  1. Regress R_{it} on X_{it} and obtain $\widehat{\zeta}_{1it}$
  2. Regress $\overset{\vee}{R}_{it}$ on X_{it} and obtain $\widehat{\zeta}_{2it}$
  3. Estimate θ_j regressing $\ln\left(\widehat{\zeta}_{jit}^2\right)$ on z_{jit}; compute $\widehat{S}_{jit} = \sqrt{\exp\left(z_{jit}\widehat{\theta}_j\right)}$, $j = 1, 2$
  4. Using $\widehat{\zeta}_{1it}$, $\widehat{\zeta}_{2it}$, $\widehat{S}_{1it}$, and $\widehat{S}_{2it}$, obtain consistent estimates via NLS:

$$\min_{\beta,\tau_1,\tau_2,\rho_1,\rho_2,\theta_\varepsilon} \sum_{i,t} \left[\begin{array}{l} \ln(FDI_{it}) - X_{it}\Pi - \tau_1 R_{it} - \tau_2 \overset{\vee}{R}_{it} \\ \quad - \rho_1 \sqrt{\exp\left(z_{\varepsilon it}\theta_\varepsilon\right)} \left(\frac{\widehat{\zeta}_{1it}}{\widehat{S}_{1it}}\right) - \rho_2 \sqrt{\exp\left(z_{\varepsilon it}\theta_\varepsilon\right)} \left(\frac{\widehat{\zeta}_{2it}}{\widehat{S}_{2it}}\right) \end{array} \right]^2$$

  5. Estimate θ_ε (again) by regressing $\ln\left(\widehat{\varepsilon}_{it}^2\right)$ on $z_{\varepsilon it}$, where

$$\widehat{\varepsilon}_{it} = \ln(FDI_{it}) - X_{it}\widehat{\Pi} - \widehat{\tau}_1 R_{it} - \widehat{\tau}_2 \overset{\vee}{R}_{it},$$

     and compute $\widehat{S}_{\varepsilon it} = \sqrt{\exp\left(z_{\varepsilon it}\widehat{\theta}_\varepsilon\right)}$
  6. Estimate via OLS:

$$\ln(FDI_{it}) = X_{it}\Pi + \tau_1 R_{it} + \tau_2 \overset{\vee}{R}_{it} + \rho_1 \left(\frac{\widehat{S}_{\varepsilon it}}{\widehat{S}_{1it}}\widehat{\zeta}_{1it}\right) + \rho_2 \left(\frac{\widehat{S}_{\varepsilon it}}{\widehat{S}_{2it}}\widehat{\zeta}_{2it}\right) + \widetilde{\varepsilon}_{it}$$

Results

KV (2009) - Type Approach

Determinants of FDI: State Effects & Time Effects

	Chemical Sector (PP&E)							
	Spec (1)		Spec (2)		Spec (3)		Spec (4)	
	CF	IV	CF	IV	CF	IV	CF	IV
ln(index)	-0.812‡	-0.558*	-0.946†	-0.218	-0.826‡	-0.419†	-0.801†	-0.384‡
	(0.462)	(0.195)	(0.476)	(0.168)	(0.424)	(0.177)	(0.388)	(0.209)
ln(spatial index)			1.482†	-0.205	-2.383	-0.978†	0.366	0.683
			(0.725)	(0.445)	(1.564)	(0.409)	(1.034)	(0.837)
Underid Test		0.000		0.000		0.000		0.000
F-stat		41.951		7.885		11.003		6.367
Overid Test		0.814		0.001		0.027		0.166
Endogeity		0.036		0.984		0.175		0.431
Joint Sign.	0.079	0.03	0.014	0.001	0.069	0.004	0.090	0.04
N	563	563	563	563	563	563	563	563

‡ p<0.10, † p<0.05, * p<0.01. Underid reports the p-value of the Kleibergen-Paap (2006) rk statistic with rejection implying identification; Overid reports the p-value of Hansen J statistic with rejection casting doubt on instruments'

Determinants of FDI: State Effects & Time Effects

	Chemical Sector (Employment)							
	Spec (1)		Spec (2)		Spec (3)		Spec (4)	
	CF	IV	CF	IV	CF	IV	CF	IV
ln(index)	-0.669†	-0.802*	-0.663†	-0.524*	-0.498	-0.563*	-0.457	-0.896*
	(0.273)	(0.135)	(0.333)	(0.131)	(0.445)	(0.124)	(0.287)	(0.168)
ln(spatial index)			-0.189	-0.143	-0.790‡	-0.771*	-1.198	-0.733
			(0.418)	(0.375)	(0.458)	(0.288)	(0.870)	(0.771)
Underid Test		0.000		0.000		0.000		0.000
F-stat		68.248		10.028		13.998		5.675
Overid Test		0.137		0.000		0.030		0.392
Endogeity		0.000		0.362		0.039		0.001
Joint Sign.	0.014	0.000	0.125	0.000	0.124	0.000	0.110	0.000
N	621	621	621	621	621	621	621	621

‡ p<0.10, † p<0.05, * p<0.01. Underid reports the p-value of the Kleibergen-Paap (2006) rk statistic with rejection implying identification; Overid reports the p-value of Hansen J statistic with rejection casting doubt on instruments' validity; Endog reports the p-value of endogeneity test of the endogenous regressors; Jsig reports the p-value of Anderson-Rubin chi-square test of endogenous regressors; F-stat reports the Kleibergen-Paap F statistic for weak identification.

Determinants of FDI: State Effects & Time Effects

	Total Manufacturing (PP&E)							
	Spec (1)		Spec (2)		Spec (3)		Spec (4)	
	CF	IV	CF	IV	CF	IV	CF	IV
ln(index)	-0.369	-0.155†	-0.472†	-0.241*	-0.115	-0.043	-0.251	-0.109
	(0.268)	(0.073)	(0.209)	(0.083)	(0.311)	(0.075)	(0.222)	(0.080)
ln(spatial index)			-0.902‡	-0.944*	0.577	-0.354‡	-0.266	0.387
			(0.490)	(0.300)	(0.368)	(0.193)	(0.572)	(0.446)
Underid Test		0.000		0.000		0.000		0.000
F-stat		137.812		10.218		20.012		8.86
Overid Test		0.758		0.000		0.001		0.211
Endogeity		0.195		0.043		0.345		0.702
Joint Sign.	0.170	0.168	0.018	0.000	0.272	0.000	0.458	0.110
N	811	811	811	811	811	811	811	811

‡ $p<0.10$, † $p<0.05$, * $p<0.01$. Underid reports the p-value of the Kleibergen-Paap (2006) rk statistic with rejection implying identification; Overid reports the p-value of Hansen J statistic with rejection casting doubt on instruments' validity; Endog reports the p-value of endogeneity test of the endogenous regressors; Jsig reports the p-value of Anderson-Rubin chi-square test of endogenous regressors; F-stat reports the Kleibergen-Paap F statistic for weak identification.

Determinants of FDI: State Effects & Time Effects

	Total Manufacturing (Employment)							
	Spec (1)		Spec (2)		Spec (3)		Spec (4)	
	CF	IV	CF	IV	CF	IV	CF	IV
ln(index)	0.042	0.042	-0.034	0.054	0.489‡	0.192*	0.160	-0.024
	(0.241)	(0.065)	(0.298)	(0.066)	(0.279)	(0.063)	(0.216)	(0.071)
ln(spatial index)			-0.310	0.131	0.323	-0.041	-0.654‡	-0.421
			(0.273)	(0.237)	(0.264)	(0.162)	(0.353)	(0.407)
Underid Test		0.000		0.000		0.000		0.000
F-stat		147.699		11.265		22.029		8.927
Overid Test		0.714		0.000		0.031		0.219
Endogeity		0.283		0.352		0.009		0.109
Joint Sign.	0.861	0.778	0.525	0.000	0.104	0.002	0.137	0.273
N	814	814	814	814	814	814	814	814

‡ p<0.10, † p<0.05, * p<0.01. Underid reports the p-value of the Kleibergen-Paap (2006) rk statistic with rejection implying identification; Overid reports the p-value of Hansen J statistic with rejection casting doubt on instruments' validity; Endog reports the p-value of endogeneity test of the endogenous regressors; Jsig reports the p-value of Anderson-Rubin chi-square test of endogenous regressors; F-stat reports the Kleibergen-Paap F statistic for weak identification.

Identification Strategy #4

Chamberlain & Griliches (1975) - Type Approach

- Basic Idea
  - Triangular system

$$
\begin{aligned}
y_i &= \beta x_i + \tau R_i + \varepsilon_i \\
R_i &= \pi x_i + \zeta_i
\end{aligned}
$$

  - Reduced forms

$$
\begin{aligned}
y_i &= (\beta + \tau\pi) x_i + [\varepsilon_i + \tau\zeta_i] = (\beta + \tau\pi) x_i + \widetilde{\varepsilon}_i \\
R_i &= \pi x_i + \zeta_i
\end{aligned}
$$

  - Note that τ is also in the reduced form error
  - Without an IV, τ is not identified either from the reduced form coefficient on x nor from the covariance matrix, Σ
  - However, we can estimate τ from Σ if we place restrictions on Σ

- With only 1 outcome, Σ is 2×2

$$\Sigma = \begin{bmatrix} \sigma_\varepsilon^2 + \tau^2\sigma_\zeta^2 + 2\tau\sigma_{\varepsilon\zeta} & \tau^2\sigma_\zeta^2 + \sigma_{\varepsilon\zeta} \\ & \sigma_\zeta^2 \end{bmatrix}$$

  which entails 3 moment conditions and 4 unknowns: $\tau, \sigma_\varepsilon^2, \sigma_\tau^2, \sigma_{\varepsilon\zeta}$
- Plausible restrictions not likely
- With additional outcomes (e.g., FDI by sector), time periods, and endogenous regressors, moments grow faster than parameters under certain (more plausible) restrictions (Chamberlain MIMIC model)

- Primary restrictions imposed
  - Endogeneity is due to a common unobserved factor, $u_{it} \sim AR(1)$
  - Factor has identical effect on all FDI measures
  - Idiosyncratic errors are *iid*
  - Measurement error in R_{it} and $\overset{\vee}{R}_{it}$ have identical variance
  - Normal distribution
- Estimation by ML imposing restrictions on Σ

$$\ln \mathcal{L}_i = -\frac{1}{2}\left[\ln|\Sigma| + \widetilde{\varepsilon}_i' \Sigma^{-1} \widetilde{\varepsilon}_i + M \ln(2\pi)\right]$$

under *iid* observations, where $\widetilde{\varepsilon}$ is a vector of reduced form errors and M is the dimension of $\widetilde{\varepsilon}_i$ $(M = 6)$

Results

Chamberlain & Griliches (1975) - Type Approach

TBD

Conclusion

- Lack of typical instruments has plagued PHH literature
- PHH literature has also ignored spatial effects known to impact FDI location
- Thus far, our atypical identification approach suggests stronger, negative effects of environmental regulation on FDI ... particularly in *pollution-intensive* industries and measured by *employment*
- Little impact of neighboring environmental regulation (but spatial effects may matter for other regressors)