A Global Database of Foreign Affiliate Activity

Tani Fukui

U.S. International Trade Commission

Csilla Lakatos

U.S. International Trade Commission

April 6, 2012

Motivation: Lack of Global Data for Foreign Affiliate Activity

Very little data on Foreign Direct investment

- U.S. BEA, UNCTAD, OECD, various national statistical agencies
- Generally produce bilateral stocks and/or flows data
- Rarely at the sectoral AND bilateral data
- CEPII (2007) does produce estimates

Even less data on Foreign Affiliate Activity

- E.g. operating activity of foreign affiliates
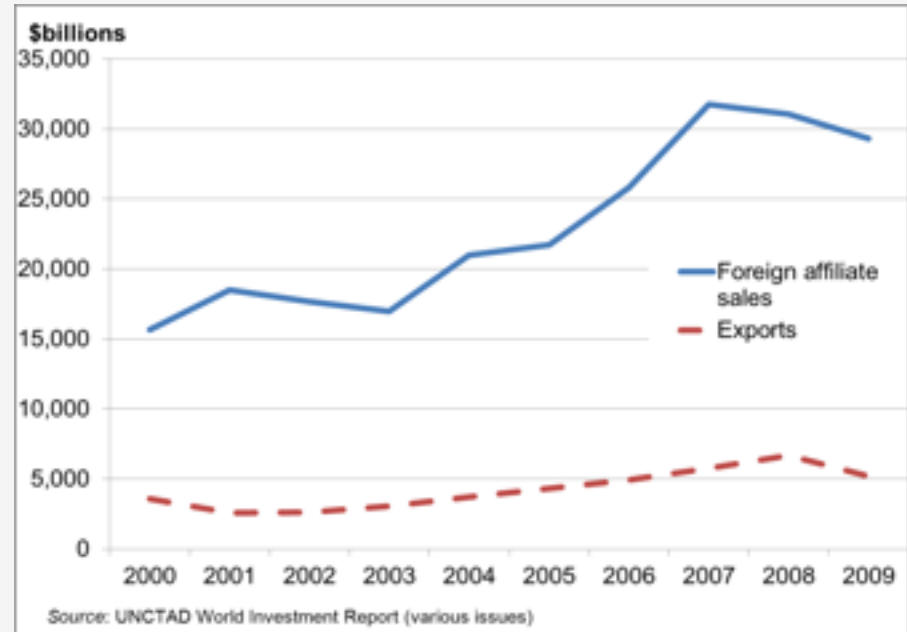- U.S. BEA, Eurostat, some national statistical agencies
- UNCTAD estimates
- Very little data, aside from these sources

Motivation: Importance of Foreign Affiliate Activity

Foreign affiliate sales are large

> Foreign affiliate sales 6 times larger than exports
>
> Estimated at $31 trillion (2007)

$billions

Source: UNCTAD World Investment Report (various issues)

- Foreign affiliate sales are growing rapidly

  7.2 percent annual growth (2000-2009)

  4 percent for exports

  2.3 percent for global economy

Source: UNCTAD World Investment Report (various issues)

Motivation: Our Work in Context

- Our Work
  - **Database:** Set of databases (1) Foreign affiliate sales; (2) FDI; (3) value added by labor and capital
  - **Model:** Global Trade Analysis Project (GTAP) model adapted to handle production by both domestic and foreign firms, and consumption of foreign and domestic products

- Prior Work
  - Hanslow (2000): limited data availability (U.S. data only), simpler modeling specification
  - CEPII FDI database and model (2007 and 2011): capital (production) side only, no sales (consumer) side
  - Other models detailing one country only

Eurostat foreign affiliate sales database

- 22 host countries
  - These are the reporters
  - EU countries

- 41 source countries
  - EU and non-EU countries
  - Mostly developed countries

- 21 sectors

- Years 2003-2007

- Sparse:
  - 48 percent missing values; 46 percent zeros
  - Remaining are positive (6,000+ observations)

Econometric Estimation

- Bergstrand and Egger (2007) and Carr, Markusen and Maskus (2001)
  - Distance and GDP measures "works" in explaining foreign affiliate sales, but why?
  - Has been done in trade: this is for FDI and FAS
- 3 factor, 3 country, 2 good model:
  - Permits the co-existence of multinational and national (domestic-only) firms
  - Incorporates third-country effects in an analogous manner to multilateral resistance terms in trade
  - Model contains both FAS and FDI variables, which generally move in tandem

Econometric Estimation

- Our adaptation: sector specific variables (production, FDI restrictiveness)

- OLS Construction:

$$\ln(FAS_{irst}) = \alpha_0 + \beta_1 \ln(\text{GDP}_{st}) + \beta_2 \ln(\text{Production}_{irt}) + \beta_3 \ln(\text{GDPROW}_{rst})$$
$$+ \beta_4 Dist + \beta_5 \ln(\text{trade openness}_{rst})$$
$$+ \beta_6 \ln(\text{investment barriers}_{rst}) + \beta_6 \ln[(S/U)_{rst}/(S/U)_{rst}] + \sum_t \gamma_t + \varepsilon_{irst}$$

  - where i = industry, r = host, s = source, and t = time

  - GDP is GDP of source country, GDP ROW is GDP of rest of world, Production is domestic production by sector of host country,

  - The skilled/unskilled variable is the difference between source and host relative skill levels

  - Dummy variables for year used

Econometric Strategies

Strategy	Description	Pros	Cons
OLS		Simplicity, common use	Usual log-log design precludes use of zero values
Poisson Pseudo Maximum Likelihood (PPML)	To address heterogeneity, Santos Silva Tenreyro (2006)	Can use and generate zero values	May not produce "enough" zeros; not suitable for overdispersed data
Zero inflated Poisson (ZIP)	Similar to the PPML estimator but with an additional zero generating "inflate"	Two ways of generating zeros may yield better results in situations with excess zeros	Assumes mean and variance to be equal. Specifying a plausible inflate process is non
Zero inflated Negative Binomial (ZINB)	Similar to ZIP but NB does not require mean equal to variance	Same as for ZINB, with the additional benefit of permitting more overdispersion in data	As with ZIP, specifying a plausible inflate process is non-trivial

Results of econometric estimation

- Consistent across all four versions

- Unexpected result: GDP of source is negative

- GDP of RoW is negative as expected

- The two sectoral variables are strongly significant and the correct sign

- Trade openness is positive indicating a positive relationship between trade and foreign affiliate sales

- Other variables as expected

	(1) OLS	(2) PPML	(3) ZIP	(4) ZINB
$Ln(GDP_{st})$	-0.0936** (-2.69)	-0.0112 (-0.41)	-0.243*** (-7.67)	-0.228*** (-5.94)
$Ln(Prod_{irt})$	0.373*** (24.77)	0.598*** (32.52)	0.456*** (21.85)	0.319*** (14.39)
$Ln(GDP\ RoW_{rst})$	-12.95*** (-21.99)	-19.07*** (-28.05)	-12.69*** (-19.07)	-12.21*** (-20.78)
$Ln(Distance_{rs})$	-0.546*** (-14.95)	-1.315*** (-26.17)	-0.652*** (-12.64)	-0.376*** (-8.05)
$Comm\ Lang_{rs}$	0.538*** (6.87)	0.288*** (3.39)	0.176* (2.08)	0.206** (2.71)
$Trad\ Open_{rt}$	0.889*** (19.37)	0.626*** (8.67)	0.783*** (10.82)	0.852*** (13.67)
$Invest\ Open_{rt}$	0.156*** (6.46)	0.0583 (1.80)	0.0836* (2.57)	0.119*** (4.21)
$FDI\ Restrict_{ir}$	-1.433*** (-9.81)	-1.267*** (-7.63)	-1.639*** (-9.29)	-1.300*** (-12.98)
$Skill\ Diff_{rst}$	1.406*** (4.53)	3.408*** (7.14)	0.722 (1.71)	1.635*** (4.57)
N	6327	43541	43541	43541
R^2	0.388	0.498		

t statistics in parentheses
* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Some Evidence Points toward ZIP/ZINB

ZIP/ZINB can handle the large amounts of excess zeros

PPML does not produce "enough" zeros, compared with the observations

ZIP/ZINB (they produce the same number of zeros) can be targeted at the desired number of zeros

Source	Positive Values	Zeros
Data	15%	85%
PPML	90%	10%
ZIP/ZINB	16%	84%

Evidence of overdispersion

- Likelihood ratio test strongly reject the assumption that the mean and variance of the underlying population are the same

Some Evidence Points to ZINB

- Examining residuals points in favor of ZIP and ZINB
- The share of fitted values that are close to the data are very high for both ZIP and ZINB
- The share of fitted values far off the data are highest for OLS

Residuals by size (share of total)
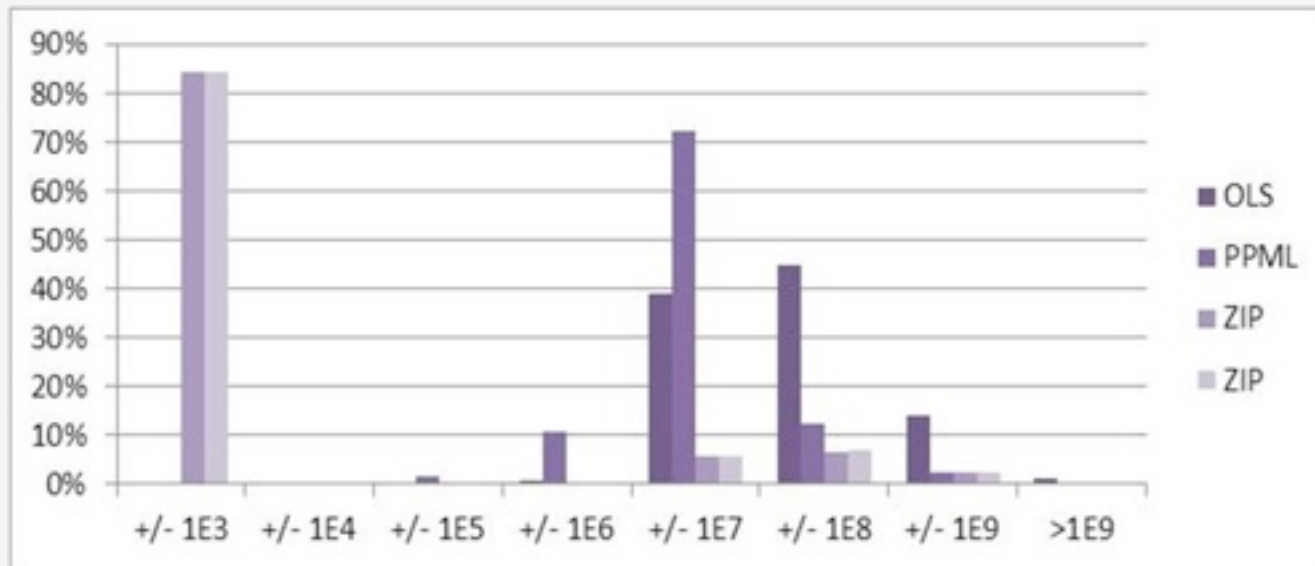
Some evidence favors OLS/PPML moderately

- Residuals of zero value observations are eliminated

- Conditional on non-zero observations, the performance across is ambiguous across strategies

Number of residuals by size (absolute values)
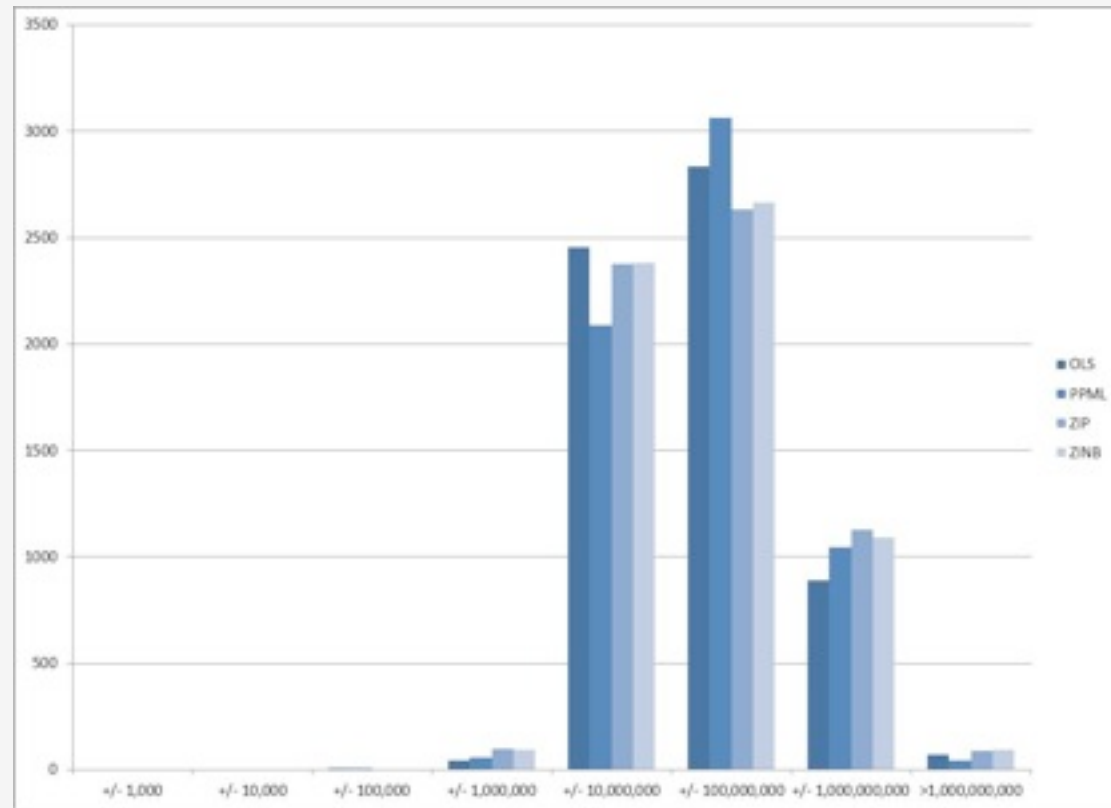
In extrapolation, ZINB underestimates

- ZINB underestimates values generally

- ZINB also predicts many zeros where data show positive values

Total Sales $(billions)	
Actual	7,313
PPML	7,121
ZINB	1,562

PPML Fitted Values Compared

ZINB Fitted Values Compared

Other ZINB problem: example

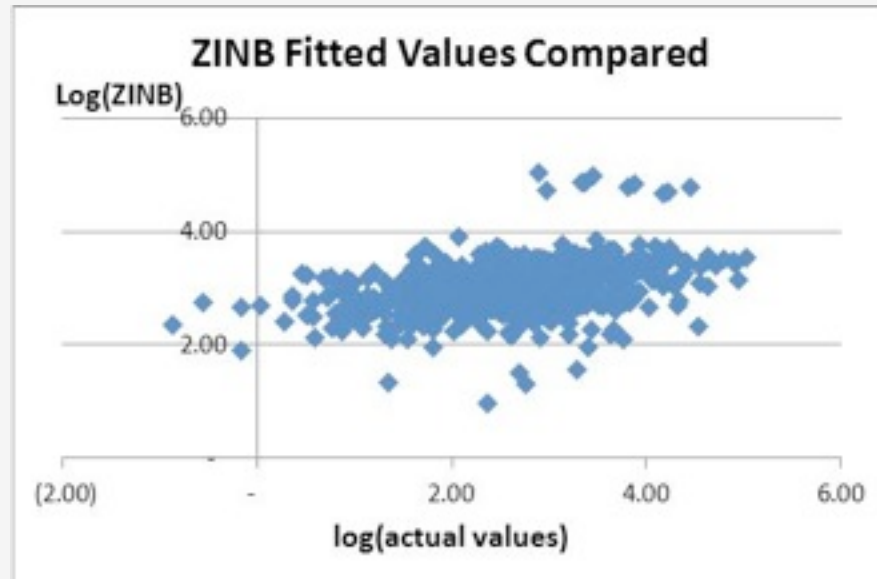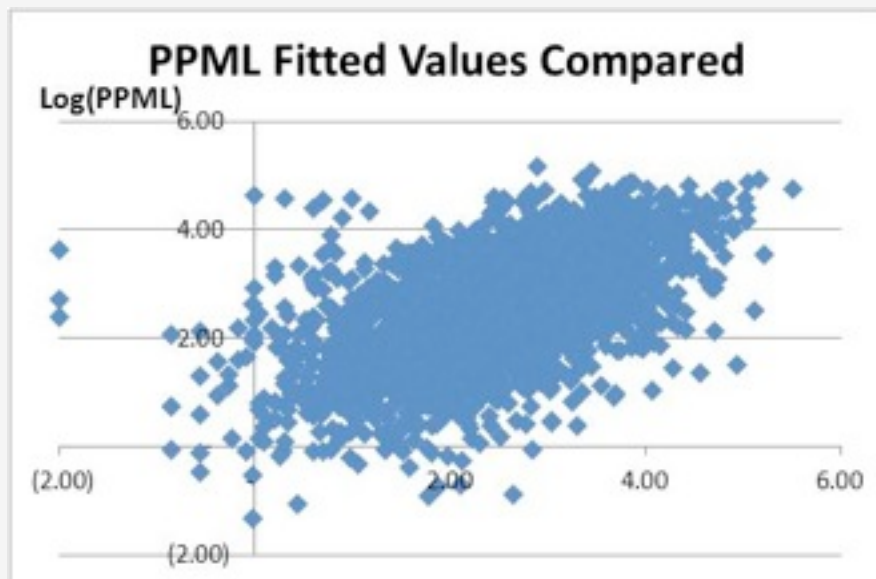| USD millions	Host countries									egy
	aut	cze	fra	deu	hun	ita	swe	bgr	rou	egy
coaoilgas										0.18
foodb_t										0.18
tex										0.14
wap										0.14
lea										0.06
lum										0.09
ppp										0.09
p_c										0.16
crp			633.97			406.49		32.3	13.69	
nmm										0.12
i_snfm										0.12
fmp										0.11
ome										
ele										0.09
mvh										0.09
otn										0.06
omf										0.06
elygdt										0.13
wtr										0.03
cns								9.85		
trd	2.6	30.8		1178.56		276.74	82.67	46.53	34.08	0.16
otp										0.16
wtp										0.08
atp										0.09
cmn										
ofi										
isr										0.09
obs	8.49		160.82	196.13	87.59	83.22				0.13

Sectors (row label on left side)

Source country is Israel

Excessive dominance of country-specific variables: e.g. Egypt is a host country in almost every sector (contiguous border variable)

Israel invests in three other sector:

- chemicals rubber and plastics ("crp"),
- wholesale/retail ("trd")
- other business services ("obs")

PPML version

USD millions	Host countries									
	aut	cze	fra	deu	hun	ita	swe	bgr	rou	egy
coaoilgas	9	5	52	-	-	-	-	-	9	77
foodb_t	18	14	244	346	14	244	23	5	23	79
tex	8	6	-	-	-	3	7	-	0	53
wap	-	-	-	-	-	-	6	-	9	54
lea	8	3	-	-	3	3	4	1	6	14
lum	8	8	82	105	0	-	7	2	9	28
ppp	18	9	156	255	8	166	26	2	7	26
p_c	18	9	145	279	9	152	20	4	13	65
crp	18	10	181	634	9	406	20	32	14	35
nmm	-	9	113	-	6	33	11	2	10	44
i_snfm	18	11	142	246	7	158	21	4	12	41
fmp	9	12	0	-	7	88	20	2	8	37
ome	9	15	179	409	-	230	28	-	12	15
ele	14	12	113	223	16	114	20	1	5	28
mvh	18	16	196	309	13	156	30	-	11	29
otn	12	6	162	98	-	108	12	-	7	17
omf	-	7	-	-	-	-	11	-	7	16
elygdt	-	11	123	-	9	-	20	5	15	50
wtr	-	4	62	-	-	-	7	1	5	5
cns	33	14	243	318	10	236	23	10	17	58
trd	3	31	265	1,179	12	277	83	47	34	66
otp	18	15	218	142	11	215	33	-	14	63
wtp	-	-	-	-	-	63	-	1	1	21
atp	-	-	2	-	4	88	0	2	-	27
cmn	-	6	-	184	7	117	15	-	10	38
ofi	16	8	168	243	6	159	16	-	6	29
isr	12	-	124	221	5	89	12	1	3	27
obs	8	19	161	196	88	83	66	5	23	48

(Sectors)

- Source country is Israel
- Now Israel invests in diverse countries and sectors with zeros scattered throughout
  - crp and trd remain dominant sectors
  - Egypt receives far less FDI from Israel (though still in every sector)

Version Used: PPML

Some observations:

- Original estimates are in col (1): we use estimates from Col (3)

- GDP source coefficient is now positive

- GDP RoW variable is more in line with other estimation strategies

- FDI restrictiveness no longer significant

- Trade openness now negative

	(1) y_round	(2) y_round	(3) y_round	y_r
$Ln(GDP_{st})$	-0.0112 (-0.41)	0.284*** (8.47)	0.280*** (8.00)	0.6 (25
$Ln(Prod_{irt})$	0.598*** (32.52)	0.479*** (25.82)	0.480*** (25.91)	0.4 (23
$Ln(GDP\ RoW_{rst})$	-19.07*** (-28.05)	-8.558*** (-13.28)	-8.612*** (-12.85)	
$Ln(Distance_{rs})$	-1.315*** (-26.17)	-0.901*** (-20.90)	-0.895*** (-22.31)	-0.7 (-2
$Comm\ Lang_{rs}$	0.288*** (3.39)	0.0304 (0.33)	0.0244 (0.25)	0. (2
$Trad\ Open_{rt}$	0.626*** (8.67)	-0.212* (-2.14)	-0.204* (-2.05)	-0. (-1
$Invest\ Open_{rt}$	0.0583 (1.80)	-0.0220 (-0.63)		
$FDI\ Restrict_{ir}$	-1.267*** (-7.63)	-0.0392 (-0.30)	-0.0453 (-0.35)	-0. (-0
$Skill\ Diff_{rst}$	3.408*** (7.14)	0.309 (0.62)	0.379 (0.73)	-3.4 (-9
$Ln(GDP_{rt})$		0.526*** (18.12)	0.526*** (18.00)	0.7 (24
$Ln(GDP/capita_{st})$		1.888*** (25.43)	1.890*** (25.42)	2.6 (28
$Ln(GDP/capita_{rt})$		0.161 (1.73)	0.145 (1.58)	-0.3 (-3
N	43541	43541	43541	43
R-sq	0.498	0.523	0.524	0.

t statistics in parentheses
* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Quadratic Optimization

- Minimizes the difference between original data (FAS0 variables) and the solution set (FAS1 variables) subject to adding up constraints that ensure consistency of database

$$\min \left\{ \begin{array}{l} \Sigma_{irs} w_{irs} \left(FAS1_{irs} - FAS0_{irs} \right)^2 + \Sigma_{rs} w_{rs} \left(FAS1_{rs} - FAS0_{rs} \right)^2 \\ + \Sigma_{ir} w_{ir} \left(FAS1_{ir} - FAS0_{ir} \right)^2 + w \left(FAS1 - FAS0 \right)^2 \end{array} \right\}$$

s.t.

$$\Sigma_{irs} FAS1_{irs} = FAS1$$

$$\Sigma_i FAS1_{irs} = FAS1_{rs}$$

$$\Sigma_s FAS1_{irs} = FAS1_{ir}$$

...

Not all constraints shown

Quadratic Optimization (cont'd)

- Data used:

  - Extrapolation based on econometrics results

  - OECD, Eurostat, U.S. BEA, China National Statistical Yearbook (manufacturing data only), UNCTAD

- We weight the data so that we can adjust for the reliability of each data source

  - Low weights mean the associated variables are permitted to vary the most (least reliable data)

Final Foreign Affiliate Sales Database: aggregation for presentation of results

- GTAP regional classification (129 regions) and 28 sectors

- For this presentation we aggregate the data into 8 regions and 5 sectors

Regions	
U.S.	ASEAN
China	A/NZ
India	EU
East Asia	ROW

Sector
Mining
Manufacturing
Wholesale/retail
Transportation
Other Services

Select targeted values

- Select targeted values ($ trillions):

	Our solution	Reported	Source (2007 data)
World	26.2	31.2	UNCTAD
EU 19 as host	6.1	5.8	Eurostat
US as host	8.1	3.6	US BEA
US as source	5.5	5.5	US BEA

- Targeted values not precisely realized in the final dataset due to the optimization technique

Sector Data Results, Compared with

- Database shows that manufacturing makes up slightly less than half of all foreign affiliate sales

- Wholesale and retail sales are significant share of foreign affiliates

- Compared with Eurostat Data
  - Global database has less manufacturing, more other services
  - Reflective of the U.S.

Sector	Sales ($bn)	Share	Shares, Eurostat
Mining	595,058	2.3%	1.8%
Manuf	11,116,264	42.5%	52.2%
Wholesale/retail	4,476,512	17.1%	19.1%
Transport	2,395,686	9.2%	7.1%
Other Services	7,584,025	29.0%	19.8%
Total	26,167,544		

Host and source country

Host	Sales ($bn)	Share
U.S.	8,121	31.0%
China	698	2.7%
India	221	0.8%
East Asia	2,159	8.2%
ASEAN	345	1.3%
A/NZ	166	0.6%
EU	10,549	40.3%
ROW	3,909	14.9%
Total	26,168	

Source	Sales ($bn)	Share
U.S.	5,511	21.1%
China	12	0.0%
India	1	0.0%
East Asia	1,711	6.5%
ASEAN	184	0.7%
A/NZ	317	1.2%
EU	14,241	54.4%
ROW	4,190	16.0%
Total	26,168	

- The EU as a whole has the greatest amount of foreign affiliate sales abroad (as source), more than twice that of the U.S.
- As host, the EU and the U.S. are closer together
- China and India are negligible sources of FAS but somewhat more significant hosts of FAS, as expected

Sector shares by host country

- Significant variation in sector investment by host country
- China has a large share of its foreign affiliate activity in manufacturing
- Australia/New Zealand is a major host of mining
- The U.S. is a host to "other services" including finance, professional services, etc.

Host	Mining	Manuf	Wholesale/retail	Transport	Other Services	Total
U.S.	0.7%	35.1%	14.1%	14.3%	35.9%	100.0%
China	0.8%	63.4%	7.6%	8.6%	19.6%	100.0%
India	2.5%	53.9%	8.6%	10.0%	24.9%	100.0%
East Asia	1.0%	51.0%	12.3%	9.0%	26.8%	100.0%
ASEAN	6.8%	28.0%	45.5%	4.5%	15.2%	100.0%
A/NZ	16.1%	33.4%	21.6%	5.3%	23.6%	100.0%
EU	1.4%	43.7%	22.4%	5.8%	26.7%	100.0%
ROW	8.0%	47.0%	11.2%	8.3%	25.4%	100.0%
Total	2.3%	42.5%	17.1%	9.2%	29.0%	100.0%

Conclusion

- Results
  - Various econometric strategies were examined: "simple" strategy worked the best
  - Produced a dataset with heterogeneity across sources, hosts, and sectors consistent with actual data

- Considerations
  - Zero inflated versions: different specifications
  - Data availability is highly skewed toward developed countries (particularly EU)
  - Better sector level explanatory variables

- Future extensions:
  - Foreign affiliates exports/imports database

Extra Slides

Actual data

	aut	cze	fra	deu	hun	ita	swe	bgr	rou	egy	uk*
trd	3	-	-	1,180	-	277	83	-	34	-	
crp	-	-	-	634	-	406	-	-	14	-	
obs	8	-	-	196	88	83	-	-	-	-	486

Zero inflate models are problematic under current specifications

- Inflate process requires a non-overlapping set of independent variables that incorporate the idea of an entry barrier
  - Possible variables: FDI restrictiveness index, Common language, Contiguous Borders, GDP per capita (source, host), GDP(source, host), distance…

- Issue
  - Most variable combinations produce a "binary" inflate process: a country either invests in every sector of a host economy or in no sector (or almost no sector) in a host economy
  - Actual data do not show this: a country typically invests in some sectors of a host country but not all
  - Alternatively, the analogous binary problem is seen with the inclusion of sector-specific variables

Introduction

- The database is part of an ongoing FDI project
- CGE model from GTAP (the Global Trade Analysis Project)
- Goal: modify this CGE model so that we can examine the effect of trade policy on FDI
- GTAP model
  - 129 regions (countries and country groupings)
  - 57 sectors (concentration in agricultural goods)
  - Requires detailed, global data to produce sensible results

Databases needed for the FDI model

- Foreign affiliate sales by source country, host country, sector – *focus of this presentation*

- FDI by source-host-sector – *adapted from an existing source*

- Value added of capital and labor by source-host-sector

Value Added

$$VAR_{irst} = \alpha_0 + \beta_1 \ln\left(GDPPC_{rt}\right) + \beta_2 \ln\left(GDPPC_{st}\right) + \delta_t + \gamma_i + \varepsilon_{irst}$$

Value added results

- Various forms of the regression turn out similar results

	(1)	(2)	(3)
GDP per capita, host	0.0645***		0.0846***
GDP per capita, source		0.0434***	0.0165***
Dummy variables?	none	none	years, sectors
R-sq	0.07	0.01	0.267
adj. R-sq	0.07	0.01	0.264

- GDP per capita of both host and source are statistically significant and positive: Positive relationship between GDP per capita and labor value added share

- Final database of value added shares will vary by host and source as well as sector.

Sector shares by source country

Somewhat less variation by source country: most countries invest near the mean (45%) in manufacturing

- Exceptions/Issues:
    - China and East Asia are weighted toward wholesale/retail sector
    - India invests heavily in manufacturing abroad

Source	Mining	Manuf	Wholesale/ retail	Transport	Other Services	Total
U.S.	3.7%	46.3%	21.6%	6.4%	22.0%	100.0%
China	0.7%	31.7%	40.5%	7.4%	19.6%	100.0%
India	1.3%	58.3%	13.6%	7.9%	18.8%	100.0%
East Asia	0.9%	40.1%	31.2%	8.0%	19.8%	100.0%
ASEAN	1.8%	47.3%	10.4%	10.5%	30.0%	100.0%
A/NZ	1.3%	45.1%	6.4%	12.0%	35.1%	100.0%
EU	1.8%	41.6%	16.2%	9.3%	31.1%	100.0%
ROW	2.6%	41.2%	9.7%	12.3%	34.2%	100.0%
Total	2.3%	42.5%	17.1%	9.2%	29.0%	100.0%

Eurostat – Some Aggregate Values

Host Country	2003	2004	2005	2006	2007
Austria	64.1	-	-	-	208.0
Bulgaria	3.4	8.7	12.5	14.8	-
Cyprus	-	0.2	0.6	1.4	1.6
Czech Republic	37.8	57.9	61.4	74.2	-
Germany	-	-	-	399.0	1,260.0
Denmark	-	-	-	37.5	77.3
Estonia	1.8	3.5	5.8	7.3	8.7
Spain	162.0	201.0	265.0	235.0	-
Finland	23.1	33.0	57.4	52.0	69.8
France	627.0	748.0	794.0	830.0	-
United Kingdom	-	-	-	994.0	1,160.0
Hungary	24.3	39.4	38.9	89.3	146.0
Italy	325.0	376.0	388.0	506.0	530.0
Lithuania	2.7	3.0	4.6	6.6	-
Latvia	2.7	3.4	5.8	8.2	11.6
Netherlands	98.8	170.0	198.0	-	287.0
Poland	-	-	-	-	188.0
Portugal	26.8	24.4	43.7	49.7	70.8
Romania	6.0	14.5	15.4	79.6	72.9
Sweden	94.1	138.0	158.0	163.0	206.0
Slovenia	3.9	4.7	-	10.0	5.2
Slovakia	12.2	19.2	22.6	20.0	29.4
Total	**1,515.6**	**1,845.0**	**2,071.7**	**3,577.6**	**4,332.3**

Context: FDI Project Map

- Database Construction

- Model

- Policy Simulation