



Foreign Affiliate Sales Data

**24th Annual Conference on Global Economic Analysis
23-25 June 2021**

Presenter: Steen Wettstein, Statistician, ERSD, WTO

Contributors: Eddy Bekkers, Erwin Corong, Indre Macskasi,
Jeanne Metivier, Wen Jin Yuan and Steen Wettstein



Overview

- **Introduction**
- **Data collection**
 - Sources
 - Adjustments
- **Data imputation**
 - Derivation
 - Mirroring
 - Backcasting, forecasting and interpolation
 - Gravity estimation
 - Cross-validation
- **Reconciling the dimensions**
 - Quadratic Optimization
- **Conclusion**



Introduction

- ▶ Foreign Affiliate Statistics
 - ▶ Inward and outward sales (or output)
 - ▶ Foreign affiliate based on the majority ownership or control
 - ▶ Classified according to their primary activity
 - ▶ Cut-off thresholds

- ▶ Objective
 - ▶ Produce a **complete and reconciled global dataset of FA sales** which can be merged with the existing GTAP database and can thus be used in computable general equilibrium (CGE)

- ▶ Final product
 - ▶ Complete and consistent FATS sales
 - ▶ 202 economies (corresponding to 141 GTAP country groups)
 - ▶ 35 industries from A to S
 - ▶ Years 2008 to 2018



Data collection – Sources

- ▶ +50 countries report some data
 - ▶ Eurostat 30 countries
 - ▶ OECD 9 countries
 - ▶ AUT, CRI, ISR, JPA, PRK, MEX, CHE, CAN & USA
 - ▶ National sources 15 countries

Number of reported observations by source

Name	Source	Inward	Outward
United States of America	NAT	4144	6927
Canada	NAT	505	1353
Zambia	NAT	304	
Israel	NAT	228	282
Serbia	NAT	220	
Thailand	NAT	120	
Viet Nam	NAT	102	
The former Yugoslav Republic of Macedonia	NAT	45	
Hong Kong, China	NAT	40	
India	NAT	33	50
China	NAT	18	20
Zimbabwe	NAT	17	
New Zealand	NAT	15	
Australia	NAT		7
Malaysia	NAT		330



Data collection - Adjustments

- ▶ **Sales** = Turnover = gross operating revenues
- ▶ **Output**=production=goods or services produced by establishment
- ▶ Measured output will be **identical** to sales for most activities, but there are **exceptions**
 - ▶ Wholesale and retail trade → sales include the value of the goods sold
 - ▶ Financial intermediation → sales only encompass services explicitly charged for
 - ▶ Insurance → sales measure total premium earned
- ▶ **Output is a superior measure** of the service supply (at least) in these cases (MSITS 2010, § 5.65)
 - ▶ *Wholesale and retail trade; repair of motor vehicles and motorcycles (G)*
 - ▶ Only **available in inward**. Outward can be derived using average shares
 - ▶ Specific treatment for inward *Financial and insurance activities (K)*
 - ▶ *K is not available in inward in Eurostat*

K64 = financial service activities, except insurance and pension funding

K65 = insurance, reinsurance and pension funding, except compulsory social security

K66 = activities auxiliary to financial service and insurance activities)

Estimation of inward K

	Division	Sales	Output
	K64	x	
+	K65		x
+	K66		x
=	K*		



Data imputation – Derivation

- ▶ **World derivation:** Derive zeros when aggregated partner world are zeros
- ▶ **Simple derivation:** Derive zeros when aggregated industries are zeros
- ▶ **FDI:** Impute zeros when FDI stocks (Eurostat, IMF, OECD, CDIS, UNCTAD, national sources) are zeros
- ▶ **Number of enterprises:** Impute zeros when the number of enterprises are zeros
- ▶ **Back/forecasting and interpolations:** using 3-year moving averages
- ▶ **Mirroring**
 - ▶ Take mirroring values if no information available in a time series
- ▶ **Discrepancies**
 - ▶ Gehlhar's Vs Shaar's method

Shaar's similarity measure:

$$S_{ij} = 1 - \text{abs}(M_{ij} - X_{ji}) / (M_{ij} + X_{ji})$$

$$Sh_{ij}^M = M_{ij} / M_i$$

$$\bar{S}_i^M = \sum_{j=1}^N Sh_{ij}^M * S_{ij}^M$$

	New Zealand- reported imports from the US (m)	US-reported exports to New Zealand (x)	Gehlhar (1996) similarity measure $ x-m /m$	proposed similarity measure $ x-m /(x+m)$
scenario 1	33	40	0.21 (not matched)	0.10
scenario 2	40	33	0.18 (matched)	0.10



Data imputation – Derivation (2)

Summary: number of data points

Description	Methodology	Inward	Outward
Reported OECD	R_OECD	8910	3430
Reported National sources	R_NAT	5783	8953
Reported Eurostat	R_EUR	694720	404194
Mirroring	M	2520622	708879
Adjustements	E7	5968	6784
Correction of negative values	E6	5	NA
Number of enterprises	E0.3	13075	7095
FDI derivation	E0.23	6354435	NA
FDI derivation	E0.22	1898	653
FDI derivation	E0.21	101138	44056
World derivation	E0.1	168980	657246
Simple derivation	E0	31547	2259446
Industries aggregated	agr	23943	25474
Total	Total	9931024	4126210



Data imputations - Gravity estimation (1)

- ▶ Impact of using mirroring value/only with reported data
 - ▶ GDP vs sectoral Value Added
 - ▶ Estimation of the missing sectoral value added
 - ▶ Pooled vs sectoral regressions using Poisson-Pseudo-Maximum-Likelihood (PPML)
 - ▶ Test specifications using out-of-sample simulations
-
- ▶ M1: $FAS_{ijst} = \exp(\alpha_1 + \beta_1 \ln(dist_{ij}) + \beta_2 col_{ij} + \beta_3 contig_{ij} + \beta_4 lang_{ij} + \beta_5 \ln(GDP_{it}) + \beta_6 \ln(GDP_{jt}) + \beta_7 \ln(GDPpc_{it}) + \beta_8 \ln(GDPpc_{jt})) * \epsilon_{ijst}$
 - ▶ M2: $FAS_{ijst} = \exp(\text{M3 coefficients} + \beta_9 DevStatus_i + \beta_{10} DevStatus_j + \beta_{11} FTA_{ijt} + \beta_{12} BIT_{ijt} + \beta_{13} WGI_{it} + \beta_{13} EU_{ijt}) * \epsilon_{ijst}$
 - ▶ M3: $FAS_{ijst} = \exp(\alpha_1 + \beta_1 \ln(dist_{ij}) + \beta_2 col_{ij} + \beta_3 contig_{ij} + \beta_4 lang_{ij} + \beta_5 \ln(VA_{ist}) + \beta_6 \ln(VA_{jst}) + \beta_7 \ln(GDPpc_{it}) + \beta_8 \ln(GDPpc_{jt})) * \epsilon_{ijst}$

M3 coefficients

Data imputations – Gravity estimation (2)



► Results of pooled regressions

	M1				M3				M2			
	Benchmark	Reported	excl.deriv	excl.mir.	Benchmark	Reported	excl.deriv	excl.mir.	Benchmark	Reported	excl.deriv	excl.mir.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
CONSTANT	-33.938*** (1.632)	-25.782*** (1.919)	-32.583*** (1.708)	-30.200*** (2.017)	-25.579*** (1.376)	-18.390*** (1.740)	-24.516*** (1.415)	-22.897*** (1.860)	-31.722*** (2.323)	-24.307*** (2.974)	-30.355*** (2.364)	-29.262*** (2.955)
LOG(DIST)	-0.456*** (0.066)	-0.487*** (0.066)	-0.439*** (0.067)	-0.498*** (0.064)	-0.477*** (0.066)	-0.508*** (0.065)	-0.458* (0.067)	-0.512*** (0.064)	-0.218* (0.122)	-0.377*** (0.109)	-0.218* (0.122)	-0.382*** (0.109)
CONTIG	0.163 (0.157)	0.186 (0.155)	0.174 (0.153)	0.224 (0.155)	0.076 (0.169)	0.143 (0.164)	0.089** (0.166)	0.176** (0.166)	0.348** (0.144)	0.372** (0.158)	0.343** (0.142)	0.407** (0.159)
COMLANG_ETHNO	0.095 (0.198)	-0.004 (0.239)	0.096 (0.191)	-0.031 (0.240)	0.078 (0.191)	-0.054 (0.235)	0.078** (0.186)	-0.084 (0.235)	0.282** (0.137)	0.147 (0.192)	0.269** (0.136)	0.129 (0.189)
COLONY	0.261 (0.214)	0.195 (0.247)	0.258* (0.209)	0.242 (0.249)	0.354* (0.198)	0.294 (0.239)	0.345 (0.194)	0.341 (0.240)	0.178 (0.132)	0.048 (0.163)	0.170 (0.133)	0.081 (0.162)
LOG(GDP_REP)	0.805*** (0.049)	0.856*** (0.057)	0.772 (0.050)	0.873 (0.057)								
LOG(GDP_PAR)	0.961*** (0.029)	0.924*** (0.038)	0.942 (0.030)	0.918 (0.038)								
LOG(VAI_REP)					0.819*** (0.046)	0.849*** (0.052)	0.787*** (0.047)	0.867*** (0.051)	0.863*** (0.040)	0.865*** (0.052)	0.833*** (0.041)	0.882*** (0.051)
LOG(VAI_PAR)					0.949*** (0.031)	0.918*** (0.042)	0.931*** (0.032)	0.903*** (0.042)	0.975*** (0.036)	0.901*** (0.041)	0.957*** (0.037)	0.885*** (0.042)
LOG(GDPPC_REP)	0.334*** (0.084)	-0.035 (0.114)	0.324*** (0.085)	-0.069 (0.105)	0.291*** (0.080)	0.053 (0.108)	0.288*** (0.081)	0.023 (0.100)	0.287** (0.110)	0.262* (0.162)	0.283** (0.113)	0.260* (0.152)
LOG(GDPPC_PAR)	1.144*** (0.088)	1.088*** (0.090)	1.096*** (0.090)	1.175*** (0.090)	1.111*** (0.090)	1.027*** (0.093)	1.064*** (0.092)	1.120*** (0.096)	1.349*** (0.131)	1.255*** (0.138)	1.287*** (0.131)	1.369*** (0.140)
DEVING_IMF_REP									0.319*** (0.116)	0.409*** (0.140)	0.321*** (0.115)	0.420*** (0.138)
DEVING_IMF_PAR									0.042 (0.129)	0.047 (0.125)	0.041 (0.128)	0.066 (0.121)
FTA_DESTA									-0.075 (0.268)	-0.539** (0.290)	-0.082 (0.272)	-0.577** (0.286)
BIT									0.369** (0.140)	0.247 (0.180)	0.327** (0.142)	0.265 (0.174)
WB_WGI									0.327** (0.130)	0.092 (0.114)	0.327** (0.131)	0.063 (0.108)
EU									0.850*** (0.128)	0.802*** (0.186)	0.804*** (0.125)	0.830*** (0.180)
Country FE	NO	NO	NO	NO	NO	NO	NO	NO				
Industry FE	YES	YES	YES	YES	YES	YES	YES	YES				
Year FE	YES	YES	YES	YES	YES	YES	YES	YES				
R2	0.491	0.53	0.497	0.519	0.536	0.569	0.538	0.558	0.591	0.633	0.587	0.625
R2 adj	0.49	0.53	0.497	0.519	0.536	0.569	0.538	0.558	0.591	0.633	0.587	0.625
# Zeros	1348929	186382	344317	338466	1348929	186382	344317	338466	1348929	186382	344317	338466
# of Mirror	1023820	0	151777	0	1023820	0	151777	0	1023820	0	151777	0



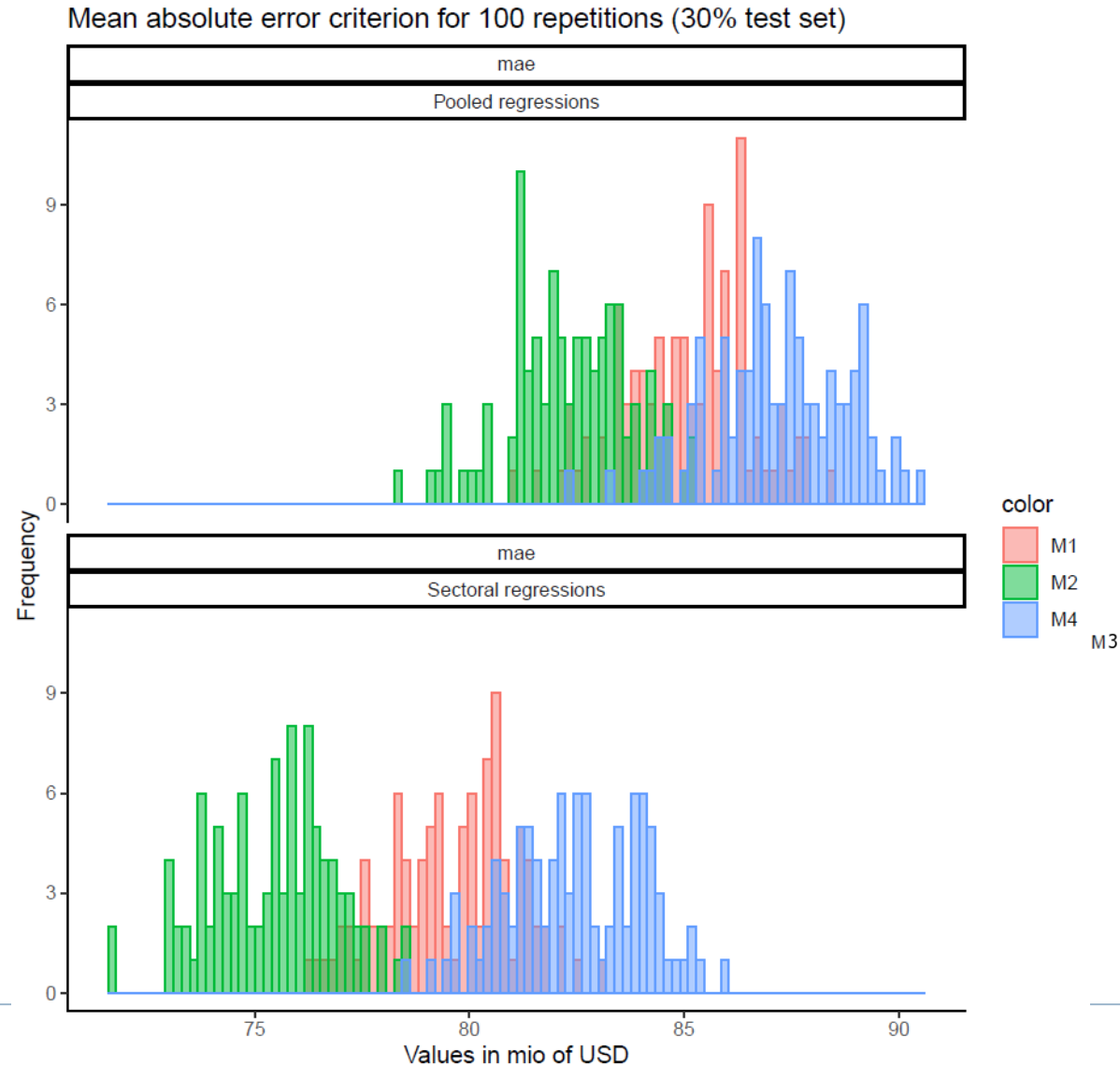
Data imputations - Gravity estimation (3)

Statistics sectoral regressions coefficients for specification 2

	-1	1	alpha = 0.05 Robust	alpha = 0.1 Robust
(Intercept)	35	0	30	30
log(dist)	28	7	20	23
contig	12	23	12	13
comlang_ethno	5	30	11	12
colony	4	31	18	19
log(VAi_rep)	0	35	35	35
log(VAi_par)	0	35	35	35
log(gdppc_rep)	13	22	10	15
log(gdppc_par)	0	35	29	29
DEVing_WTO_rep	17	18	11	15
DEVing_WTO_par	26	9	13	14
FTA_DESTA	14	21	7	9
BIT	15	20	11	13
WB_WGI	16	19	13	14
EU	5	30	18	24



Data imputation – cross validation (1)





Data imputation – cross validation (2)

Means of sectoral Means Absolute Errors (MAE) for 100 repetitions

Name	Industry	Specifications		
		M1	M2	M3
Agriculture, forestry and fishing	A	6.3	5.0	7.0
Mining and quarrying	B	75.7	74.2	93.2
Extraction of crude petroleum and natur	B06_B09	5.5	6.6	11.8
Other manufacturing	C_OTH	53.8	54.4	59.8
Manufacture of food products; ...	C10-C12	93.3	93.0	88.7
Manufacture of textiles ...	C13_C14	2.2	1.9	1.9
Manufacture of wood, paper...	C16-C18	20.8	20.4	19.9
Manufacture of coke and ...	C19	29.2	34.1	32.2
Manufacture of chemicals ...	C20	82.0	80.1	80.9
Manufacture of basic pharma...	C21	41.4	39.6	46.8
Manufacture of rubber and plastic...	C22	40.7	40.4	43.4
Manufacture of basic metals and fabr...	C24_C25	85.9	81.5	82.0
Manufacture of computer, electronic...	C26	132.2	115.5	132.6
Manufacture of machinery and equipme	C28	92.8	93.0	99.9
Manufacture of motor vehicles, trail...	C29	150.3	141.1	159.5
Manufacture of other transport equipm	C30	21.2	20.8	20.0
Electricity, gas, steam and air c...	D35	102.4	102.8	103.1
Water supply; sewerage, waste manag...	E	10.4	10.2	10.3

Name	Industry	Specifications		
		M1	M2	M3
Construction	F	45.3	44.3	46.0
Wholesale and retail trade; repair...	G	1,151.1	1,067.2	1,194.9
Land transport and transport via ...	H49	13.8	13.4	13.9
Water transport	H50	3.6	3.5	3.6
Air transport	H51	4.1	3.8	4.1
Warehousing and support activities...	H52	34.5	33.6	34.3
Postal and courier activities	H53	1.8	1.6	1.7
Accommodation and food service..	I	18.6	18.0	18.5
Information and communication	J	169.6	163.0	177.3
Financial service activities, except ins..	K64	32.3	27.8	34.3
Insurance, reinsurance and pension...	K65	127.0	110.7	120.2
Real estate activities	L	19.6	18.4	19.7
Professional, scientific and technical...	M	109.5	107.0	113.7
Administrative and support service...	N	62.8	61.6	64.5
Education	P85	0.6	0.6	0.6
Human health and social work activitie	Q	3.5	3.4	3.5
Arts, entertainment and recreation	R	2.5	2.5	2.6
Other service activities	S	2.9	2.9	2.8





Results before the reconciliation

Table: Summary of the bilateral FATS dataset before reconciliation

	Description	data points	% data points	value	% value
Reported	Reported OECD	5,217	0.0%	142,634	0.1%
	Reported National sources	1,125	0.0%	5,592,365	2.4%
	Reported Eurostat	212,845	1.4%	42,339,873	18%
	Mirroring reported	142,638	0.9%	22,901,355	10%
	Mirroring	826,907	5.3%	4,143,639	2%
Estimated	Simple derivation	128,634	0.8%	-	0.0%
	PPML predictions	13,343,573	86%	118,362,481	51%
	Back-forecasting, interpolations	834,479	5.4%	31,475,449	14%
	Correction of negative values ₁		0.0%	-	0.0%
	Industries aggregated	21,725	0.1%	3,290,982	1.4%
	Adjustements	547	0.0%	3,671,180	1.6%
Total reported (+ mirror)		1,188,732	8%	75,119,866	32%
Total estimated		14,328,959	92%	156,800,092	68%

Reconciling the dimensions – Quadratic Optimization



Minimize

$$\begin{aligned} & \sum_{c,r,s} w_{c,r,s} \left(FATS_{c,r,s}^1 - FATS_{c,r,s}^0 \right)^2 + \sum_{r,s} w_{r,s} \left(FATS_{r,s}^1 - FATS_{r,s}^0 \right)^2 \\ & + \sum_{c,r} w_{c,r} \left(FATS_{c,r}^1 - FATS_{c,r}^0 \right)^2 + \sum_{c,s} w_{c,s} \left(FATS_{c,s}^1 - FATS_{c,s}^0 \right)^2 \\ & + \sum_r w_r \left(FATS_r^1 - FATS_r^0 \right)^2 + \sum_s w_s \left(FATS_s^1 - FATS_s^0 \right)^2 \end{aligned}$$

subject to

$$\sum_{c,r,s} FATS_{c,r,s}^1 = FATS_{GLOBAL}^{Data}$$

$$\sum_c FATS_{c,r,s}^1 = FATS_{r,s}^{Data}$$

$$\sum_s FATS_{c,r,s}^1 = FATS_{c,r}^{Data}$$

$$\sum_r FATS_{c,r,s}^1 = FATS_{c,s}^{Data}$$

$$\sum_{c,s} FATS_{c,r,s}^1 = FATS_r^{Data}$$

$$\sum_{c,r} FATS_{c,r,s}^1 = FATS_s^{Data}$$



Conclusion

- ▶ Ongoing work
- ▶ Build upon previous work
- ▶ Time invested in the modeling
- ▶ Improvement of the methodology
- ▶ The complete FATS sales data will be soon available