

# Microdata Infrastructure (MDI)

Eric Bartelsman and the MDI Team\*

Vrije Universiteit Amsterdam and Tinbergen Institute

\*This project is partly funded by the European Union via the Technical Support Instrument



CompNet The Competitiveness Research Network

## (Cross-country) research with firm-level data is difficult

- Commercial sources are expensive and limited in use (linkability, quality)
- Getting access to National Statistical Institute (NSI) data not always as straightforward (... as at CBS).
- Startup costs of researcher at NSI are high: documentation and data are 'messy'
- Confidential data cannot be combined outside NSI (e.g. across sources or countries)

### MDI helps:

- **Metadata** for firm-level files and variables over time in each country
- **Harmonization** and documentation of linked longitudinal panels (MD datasets)
- **R-package MDI** with tools to read meta-data, harmonize datasets and aid writing research code
- **Mock data** of NSI datasets and of harmonized MD panels, to write and test code
- **Launches** of researchers' code in multiple countries. Output can be combined across countries

### Harmonized MD Datasets



### Metadata: NSI Variables

Example: Metadata for NSI Bus Stat variables in Netherlands

NSI_datafile	NSI_varname	is_key	description	class	domain
sbs2018	ent_id	1	Enterprise ID	character	**
sbs2018	sbs_12110	0	Turnover	numeric	1000
sbs2018	sbs_12150	0	Value added	numeric	1000
sbs2018	sbs_13310	0	Personnel costs	numeric	1000
sbs2018	sbs_16130	0	Number of employees	numeric	1

### Metadata: Harmonization

Example: Metadata to create harmonized MD Balance Sheet variables from Slovenian data

NSI_filename	NSI_varname	MD_varname	method	detail
ZR_GOSPODARSTVO_2_2004	MS_7_razST	firmid	remap	
ZR_GOSPODARSTVO_2_2004	AOP017_TEK_L	fixed_assets	revalue	$x \times 239.64/1000$
ZR_GOSPODARSTVO_2_2004	AOP030_TEK_L	fixed_assets	revalue	$x \times 239.64/1000$
ZR_GOSPODARSTVO_2_2004	AOP017_TEK_L	fixed_assets	redefine	+
ZR_GOSPODARSTVO_2_2004	AOP030_TEK_L	fixed_assets	redefine	+

### Harmonization Operations

- **revalue** Values changed using R function
- **recode** Values changed using a codebook
- **remap** Variable name is changed
- **redefine** MD variable is function of multiple NSI variables.

## MDI 'Launch'

### Initialization

- Disclosure Rules
- Directory Paths

### Harmonization

- Read MDI Tools and Metadata
- Map Variables to MD Datasets

### Research

- Run Research Code
- Prepare Output

### Cross-country

- Combine Output
- Analyze and Report

Currently running as a pilot in AT, DE, FI, FR, NL, SI, PT

## Example: Heterogeneous Productivity and the Phillips Curve

- Authors: D. Aglio and E. Bartelsman
- Question: How does price respond to demand and cost shocks?
- Countries: FR, NL, SI, (FI)
- MD datasets: BR, SBS, Prodcum, ITGS
- External data: Comtrade, WIOD

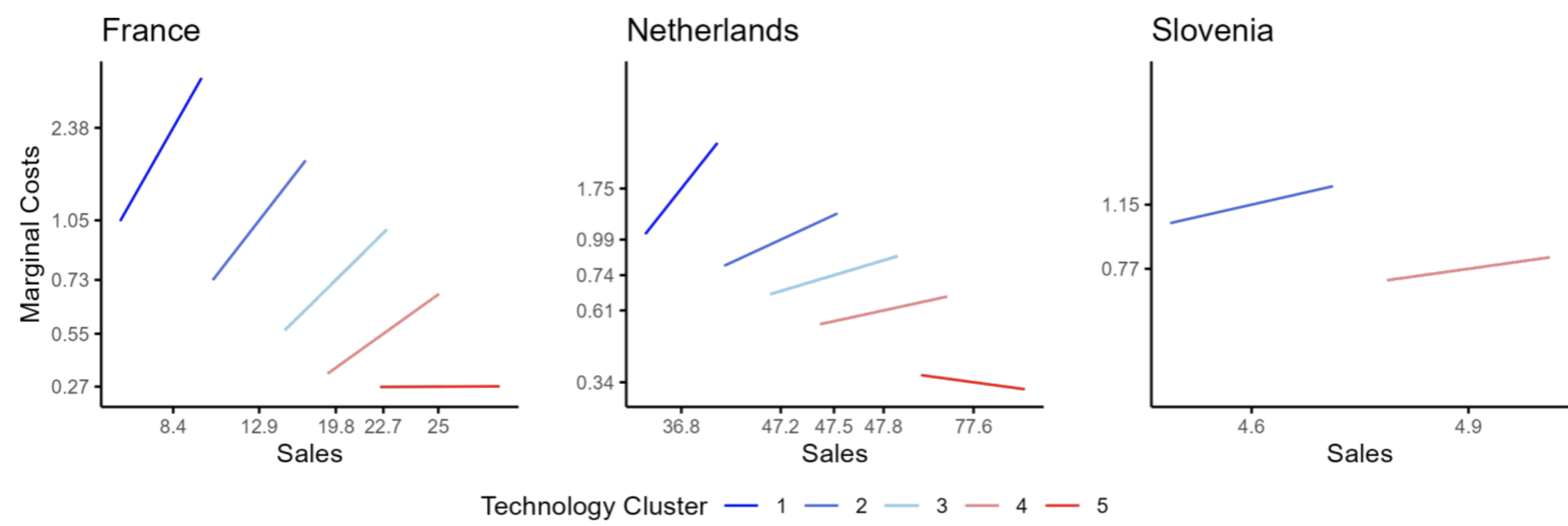


Figure 1: Supply curves by technology cluster

## Example: Firm Responsiveness over the Business Cycle

- Authors: C. Yi, A. Ferreira, J. Miranda, L. Indraccolo, E. Weder
- Question: How do comparable firms adjust to productivity shocks?
- Countries: FR, NL, PT, SI
- MD datasets: BR, BS, SBS
- External data: Ameco

