



Dipartimento
delle Finanze

BIG-DF: I “Big data” del Dipartimento delle Finanze

Roma, 23 Maggio 2017



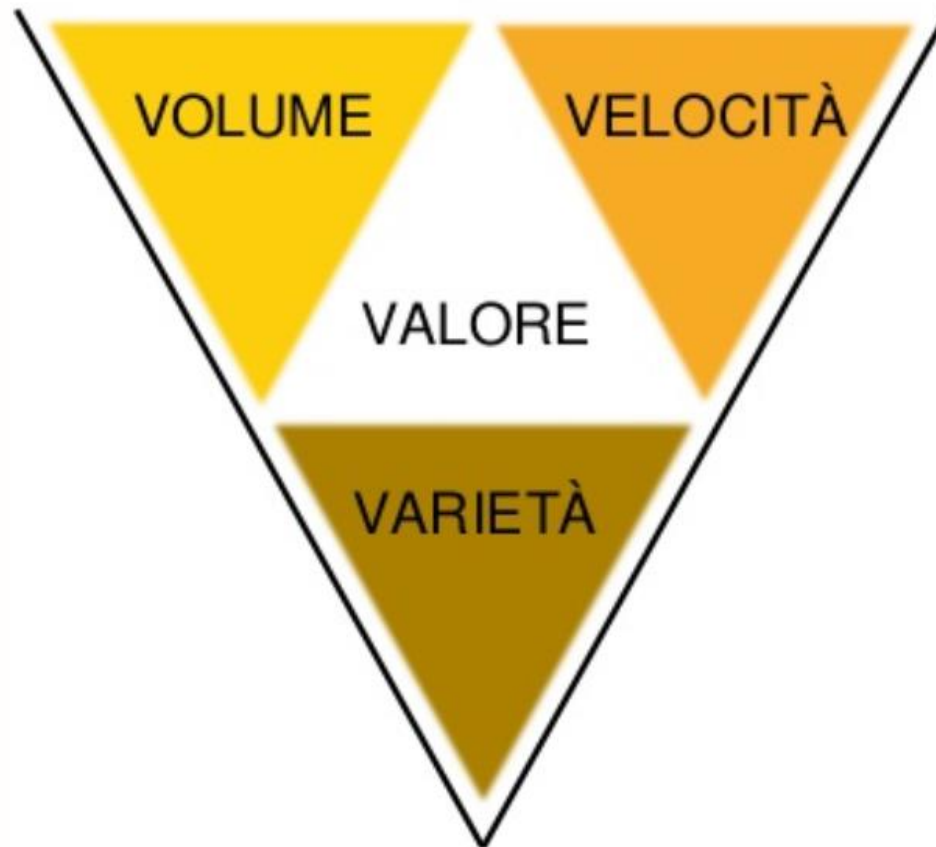
Sommario

- Big-Data: Definizione e Caratteristiche
- BIG-DF -Il Processo di Integrazione dei Dati-: una Nuova Sfida per il Dipartimento delle Finanze
- L'Evolutione dei Modelli: dai Dati Campionari ai Dati Panel di Fonti Amministrativa
- Utilizzo del BIG-DF e Applicazioni Pratiche
- Sviluppi Futuri

Big-Data: Definizione e Caratteristiche

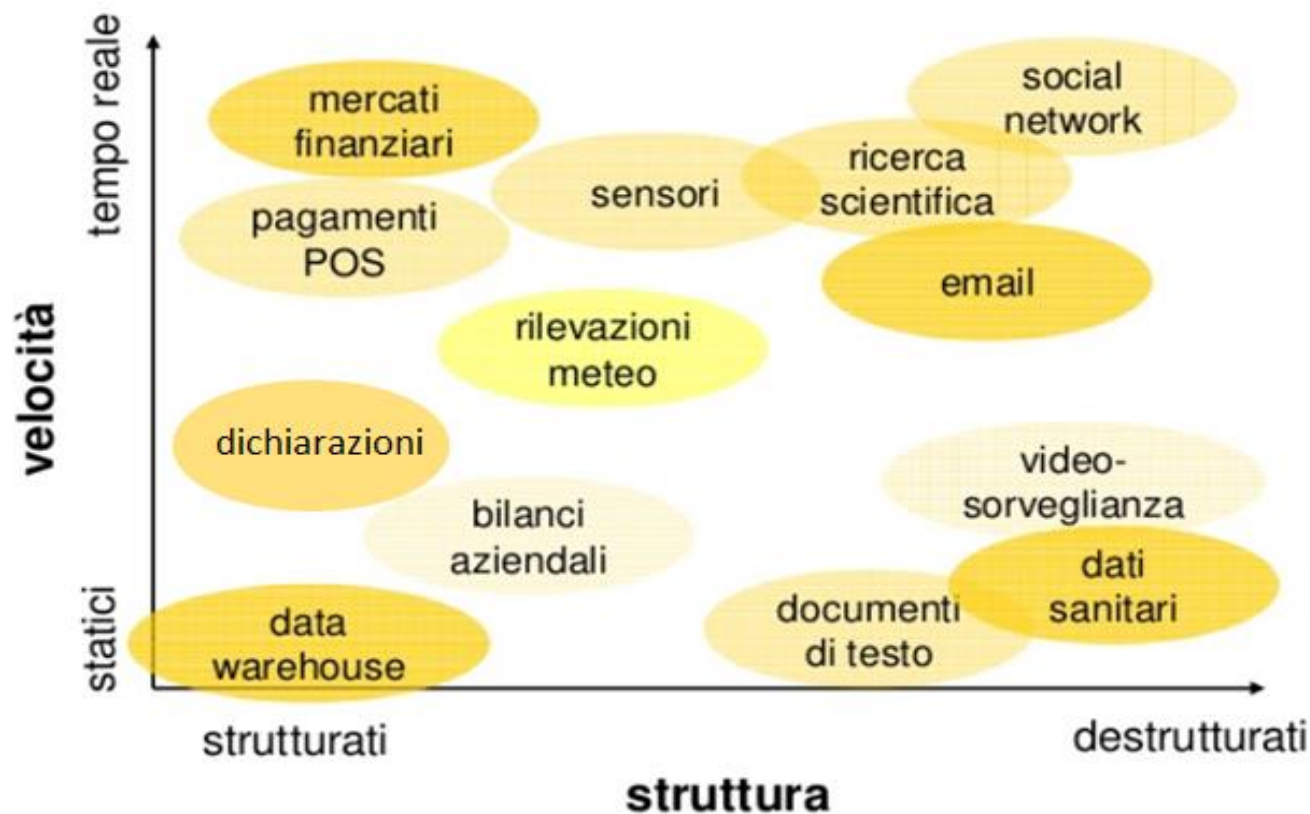
- Col termine **Big Data** *“intendiamo una raccolta di dati così estesa in termini di volume, velocità e varietà da richiedere tecnologie e metodi analitici specifici per l'estrazione di valore”*. (De Mauro et al. 2016)
- 4 temi dei Big Data: Informazione (Laney's «3 Vs»: Volume, Velocity, Variety); Tecnologia; Metodi; Impatto
- Nuove sfide per l'analisi statistica, per estrarre valore dai big data, per l'elaborazione dei dati.
- Necessità di nuove competenze e capitale umano, non solo dal punto di vista economico e statistico

Big-Data: Definizione e Caratteristiche



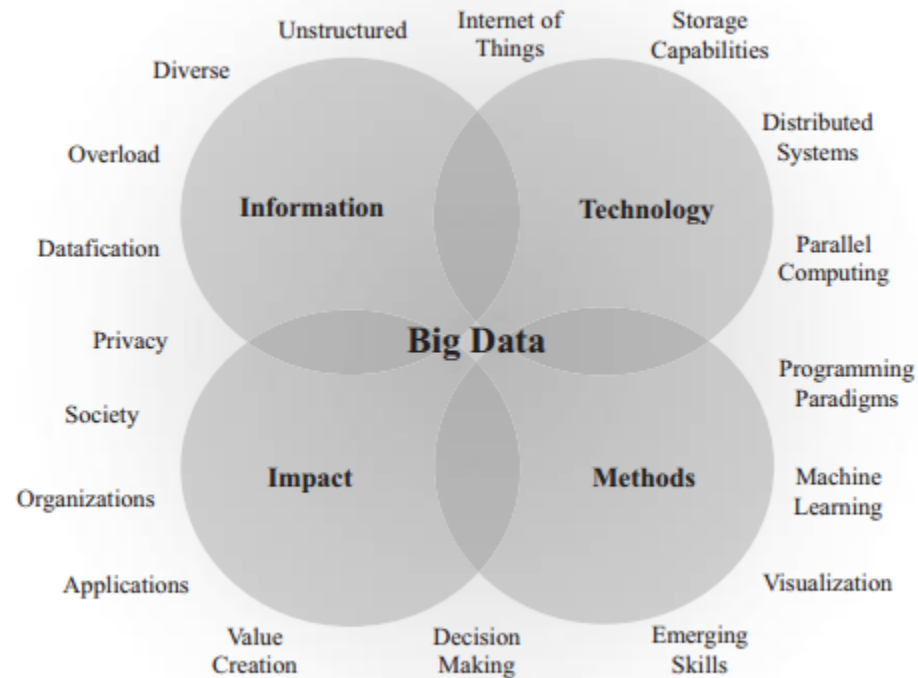
Fonte: Rapporto MIUR 2016. Vedi: Laney (2001), Beyer e Laney (2012), Zikopoulos e Eaton (2011), Zaslavsky et al. (2013). Vedi anche: Schroeck et al (2012), Dijcks (2013)

Caratteristiche dei Big Data: velocità e varietà



Cfr.: Rapporto MIUR 2016

Big-Data: Definizione e Caratteristiche



Fonte: De Mauro, Greco e Grimaldi 2016

Big-Data: Definizione e Caratteristiche

Vantaggi:

- ampliano le opportunità di analisi;
- aumentano la tempestività delle informazioni;
- possono contribuire a migliorare la qualità delle stime.

Svantaggi:

- mancanza di adeguate professionalità in grado di analizzare e interpretare i dati;
- difficoltà nella elaborazione e nell'analisi dei dati (problemi tecnologici dovuti al trattamento di ingenti quantità di dati)
- disponibilità dei dati: non sempre le informazioni sono disponibili o sia per problemi di riservatezza sia per assenza dei dati stessi;
- *privacy*; il nostro codice della privacy, il decreto legislativo 196 che riflette strettamente i contenuti della direttiva europea 95/46, risulta particolarmente deficitario rispetto alla sfida lanciata dai Big data.

BIG-DF -Il Processo di Integrazione dei Dati-

- BIG-DF è un'infrastruttura informatica costruita dal Dipartimento delle Finanze, con la partnership tecnologica di SOSE s.p.a.
- Il collegamento al server consente di elaborare dati derivanti da differenti fonti amministrative, anche attraverso l'utilizzo di software statistici (SAS, Stata)
- La banca dati delle società di capitali (fonte AIDA dei bilanci delle società di capitali) contiene le seguenti informazioni:
 - ✓ **Conto Economico;**
 - ✓ **Stato Patrimoniale;**
 - ✓ **Indicatori di gestione.**

BIG-DF -Il Processo di Integrazione dei Dati-

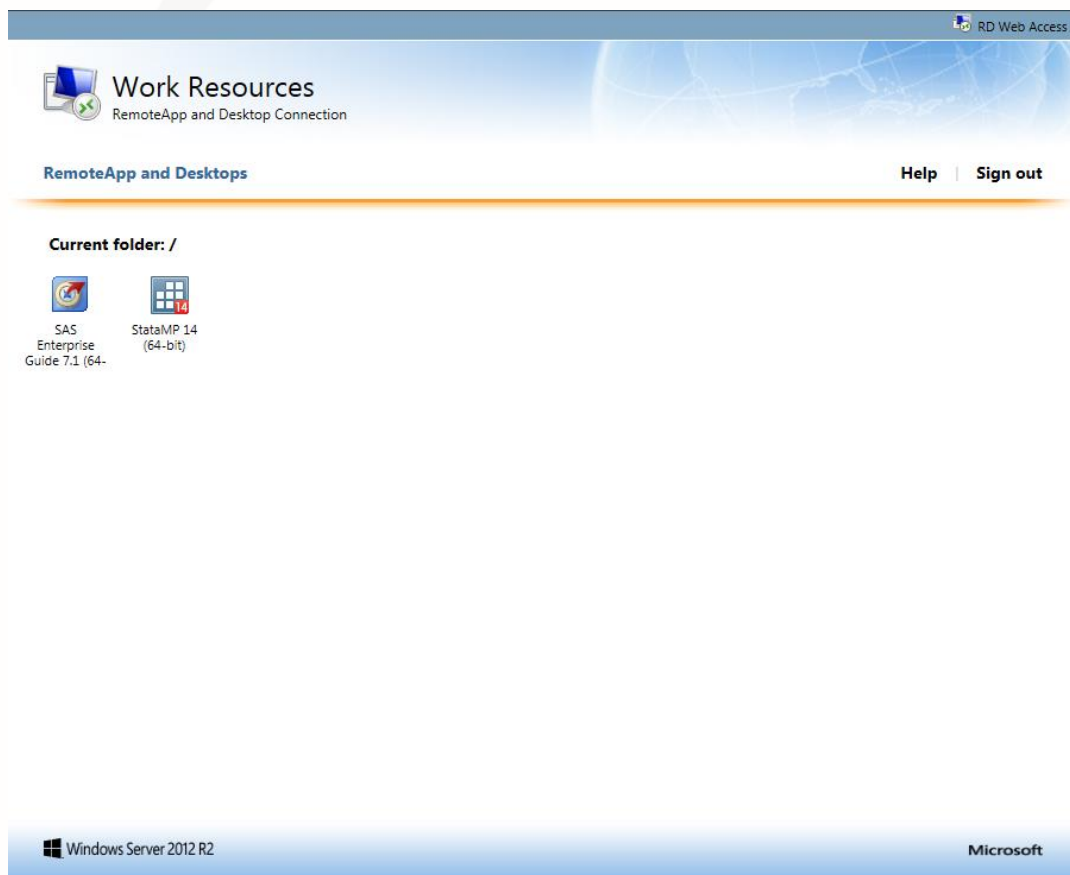
La banca dati delle società di capitali è arricchita e integrata con le informazioni già a disposizione del Dipartimento delle Finanze:

- ✓ **dichiarazioni dei redditi (Unico, IVA, IRAP, 770 ...);**
- ✓ **spesometro (informazioni sui clienti/fornitori);**
- ✓ **fatturazione elettronica;**
- ✓ **pagamenti delle imposte (F24);**
- ✓ **archivio degli immobili.**

BIG-DF -Il Processo di Integrazione dei Dati-

- In questo modo il Dipartimento delle Finanze si dota di una banca dati unica nel suo genere per la ricchezza delle informazioni raccolte dai dati disponibili in Anagrafe Tributaria.
- Tale banca dati costituisce il Big Data del mondo delle imprese italiane su Dati Fiscali.
- Tale strumento permette di gestire, in termini di memoria, il panel di imprese dal 2003 al 2015 (oltre 10 milioni di osservazioni).

BIG-DF -Il Processo di Integrazione dei Dati-



BIG-DF -Il Processo di Integrazione dei Dati-

The screenshot displays the Stata/MP 14.2 software interface. The main window shows the 'Review' tab with the following text:

```
10-user 16-core Stata network perpetual license:  
Serial number: 501406330636  
Licensed to: Emanuele Schirru  
SOSE S.p.A.
```

Notes:

1. Unicode is supported: see [help unicode_a](#)
2. More than 2 billion observations are all
3. Maximum number of variables is set to 50
4. New update available: type `-update all-`

running C:\Program Files (x86)\Stata14\profile.do

Checking for updates...
(contacting <http://www.stata.com>)

Update status
Last check for updates: 09 May 2017
New update available: 04 May 2017 ([what's](#))
Current update level: 14 Sep 2016 ([what's](#))

Possible actions
[Install available updates](#) (or type

Click to [edit automatic update checking preferences](#)

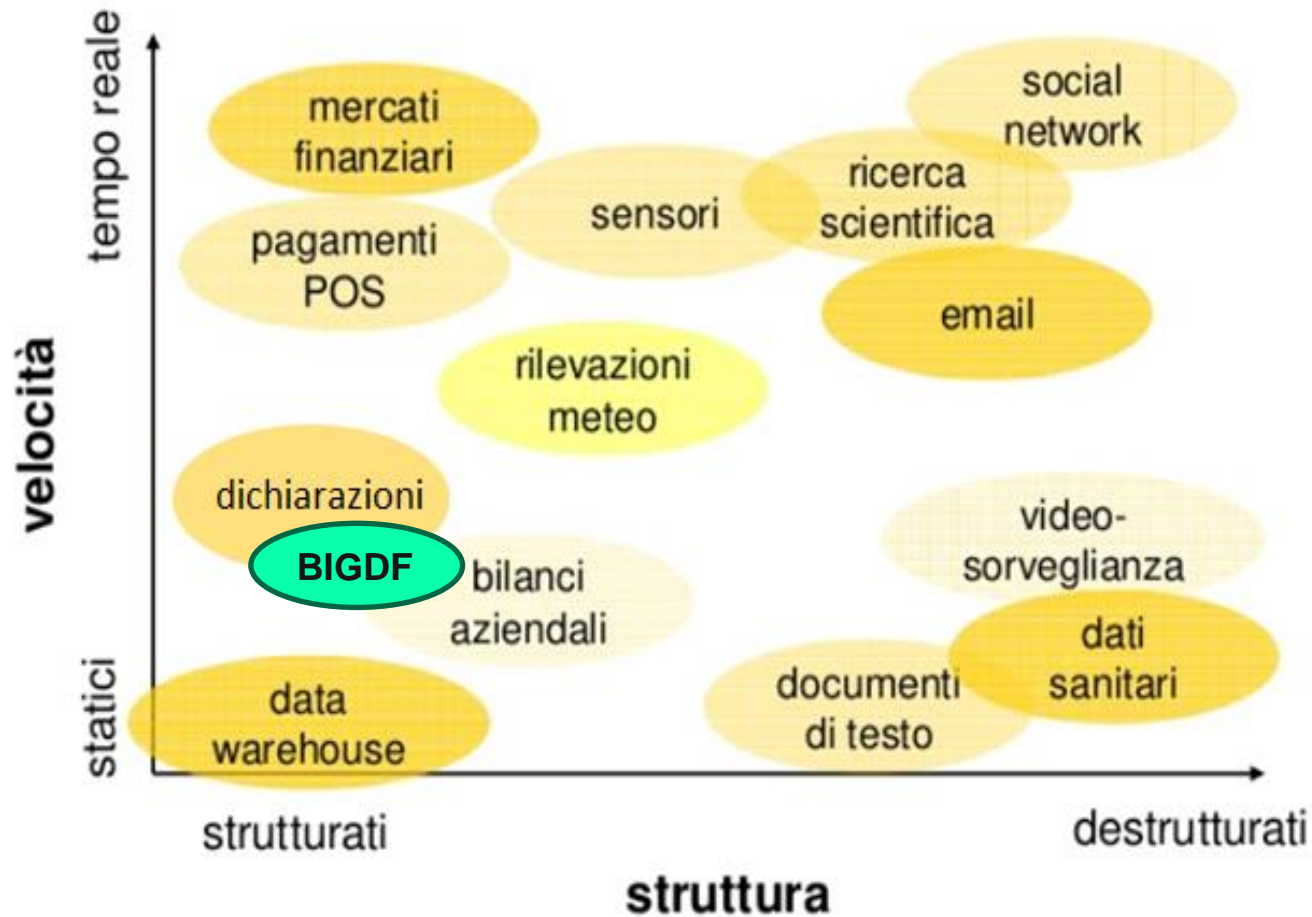
Command

The 'Open' dialog box is open, showing the file explorer for 'DATA (D:) \ BIGDF'. The file list is as follows:

Name	Date modified	Type
BIGDF_2004_2014	24/02/2017 10:15	Stata Dataset
BIL_INDISTR_2014	28/02/2017 16:03	Stata Dataset
cf_bil_04_14_in_att	28/02/2017 16:47	Stata Dataset
IRAP_SC_AI_2014	28/02/2017 14:47	Stata Dataset
IVA_SC_AI_2014	28/02/2017 14:51	Stata Dataset
UNICO_SC_AI_2004	03/03/2017 12:34	Stata Dataset
UNICO_SC_AI_2005	03/03/2017 12:33	Stata Dataset
UNICO_SC_AI_2006	03/03/2017 13:04	Stata Dataset
UNICO_SC_AI_2007	03/03/2017 13:05	Stata Dataset
UNICO_SC_AI_2008	03/03/2017 13:27	Stata Dataset
UNICO_SC_AI_2009	03/03/2017 13:32	Stata Dataset
UNICO_SC_AI_2010	03/03/2017 14:24	Stata Dataset

The 'File name' field is empty, and the file type is set to 'Stata Data Files (*.dta)'. The 'Open' and 'Cancel' buttons are visible.

BIG-DF -Il Processo di Integrazione dei Dati-



BIG-DF: una Nuova Sfida per il Dipartimento delle Finanze

- 1. funzione conoscitiva** → individuazione di anomalie nelle dichiarazioni dei redditi anche ai fini della prevenzione dei comportamenti anomali da parte dei contribuenti
- 2. funzione valutativa** → assicurare la valutazione degli effetti finanziari sia ex ante sia ex post;
- 3. funzione di analisi** → analisi comportamentale delle scelte delle imprese

BIG-DF una Nuova Sfida per il Dipartimento delle Finanze

- Prospettive di integrazione e uso di nuove fonti di dati
- Necessità di archiviare e analizzare quantità sempre maggiori di dati
- Confrontare i modelli basati su dati campionari e i nuovi modelli di microsimulazione basati sui dati dell'intera popolazione
- «Da un modello «tradizionale», basato sull'acquisizione diretta dei dati, ad un modello basato sull'utilizzo dei registri statistici, essenzialmente derivati dalle fonti amministrative e alimentati nel continuo da flussi telematici» (Alleva, 2015)

L'Evoluzione dei Modelli: dai Dati Campionari ai Dati Panel di Fonti Amministrativa

- L'utilizzo di dati non campionari implica una evoluzione della modellistica
 - Modelli di analisi ex-ante (modelli di microsimulazione e analisi comportamentale, analisi predittiva, ecc...)
 - Modelli di valutazione ex-post (inferenza causale)
- L'amministrazione finanziaria deve investire in competenze;
- È necessario attivare dei percorsi formativi per analizzare e interpretare i dati;
- Non sempre sono necessari semplici competenze informatiche e/o statistiche; spesso occorrono economisti ma anche data scientist.

BIG-DF : applicazioni pratiche

- Analisi degli effetti dell'introduzione dello *Split Payment*
- Analisi degli effetti delle nuove regole del *Reverse Charge*
- Analisi degli effetti dell'introduzione della *Patent Box*
- Analisi economica delle performance delle *start-up* innovative
- Analisi economica delle performance delle PMI innovative
- Analisi economica delle performance delle imprese esportatrici
- Analisi economica delle performance delle imprese residenti con controllate all'estero
- Analisi economica delle performance delle imprese nazionali con controllo estero

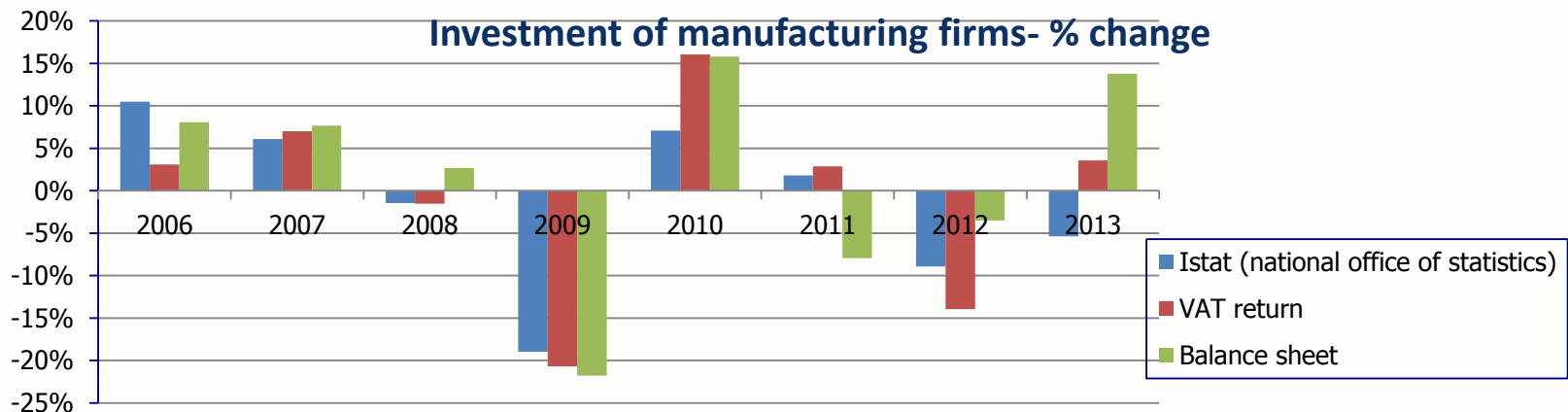
BIG-DF: La dinamica degli investimenti in Italia

- Il *dataset* BIG-DF ha permesso di implementare un modello finalizzato ad analizzare le determinanti degli investimenti, stimando diverse specificazioni di funzioni dell'investimento comunemente utilizzate nella letteratura economica ed empirica.
- In particolare, si è approfondito il ruolo e gli effetti della politica fiscale sugli investimenti utilizzando una serie di variabili fiscali e di bilancio di imprese operanti nel settore manifatturiero nel periodo 2003-2015.

BIG-DF: La dinamica degli investimenti in Italia

Per costruire la variabile relativa agli investimenti sono state utilizzati due metodi:

1. inferire gli investimenti dai dati di bilancio, utilizzando la seguente formula: $I_t = K_t - K_{t-1} + \delta K_t$;
2. utilizzare la variabile "acquisto di beni strumentali" dalle dichiarazioni IVA.



BIG-DF: La dinamica degli investimenti in Italia

Per misurare le determinanti sono state utilizzate due tipi di specificazione del modello:

1. modello dell'equazione di Eulero:

$$\frac{I_{i,t}}{K_{i,t-1}} = \beta_1 \frac{I_{i,t-1}}{K_{i,t-2}} + \beta_2 METR_{i,t-1} + X'_{it} \beta_3 + a_i + d_t + \varepsilon_{it} ;$$

2. si basa su un modello a correzione d'errore (*ECM o error correction model*). In particolare, è stata stimata la seguente equazione:

$$\frac{I_{i,t}}{K_{i,t-1}} = \beta_1 \frac{I_{i,t-1}}{K_{i,t-2}} + \beta_2 TAUC_{it} + \beta_3 \Delta y_{it} + \beta_4 \Delta y_{i,t-1} + \beta_5 \frac{CF_{i,t}}{K_{i,t-1}} + \beta_6 \frac{CF_{i,t-1}}{K_{i,t-2}} + \beta_7 (k_{i,t-2} - y_{i,t-2}) + a_i + d_t + \varepsilon_{it}$$

- Tale analisi è stata utilizzata per la valutazione ex-ante della misura del maxi-ammortamento.
- L'aumento degli investimenti stimato, pari a 3,8 miliardi di euro, è stato in linea con i risultati raggiunti nel 2016.

BIG-DF: Sviluppi Futuri

- Modello di microsimulazione IRES (dati Dichiarazioni e Bilanci) e IRAP
- Sviluppi futuri eventuali: integrazione dati con i bilanci delle società di capitali e le catene di controllo (fonte Orbis).
- Indicatori di pressione fiscale (aliquote marginali e medie effettive *backward looking* e *forward looking*)
- Aumentare la tempestività nella elaborazione dei dati.

BIG-DF: La dinamica degli investimenti in Italia

Il modello dell'equazione di Eulero

	(1)	(2)	(3)	(4)	(5)
Dependent variable: I_t/K_{t-1}					
I_{t-1}/K_{t-2}	0.0911*** (0.0100)	0.0950*** (0.0100)	0.0794*** (0.0125)	0.0887*** (0.00995)	0.0861*** (0.0101)
$METR_{t-1}$		-0.108** (0.0527)	-0.130** (0.0555)	-0.262*** (0.0529)	-0.256*** (0.0535)
$(\text{Equity/Total liabilities})_{t-1}$		0.459*** (0.0390)	0.450*** (0.0401)	0.427*** (0.0383)	0.422*** (0.0389)
$\text{Cash Flow}_{t-1}/K_{t-2}$			0.0204** (0.00991)		
ROI_{t-1}				0.507*** (0.0522)	0.507*** (0.0529)
$\text{Orders (by subsector)}_{t-1}$					0.000774*** (0.000182)
Observations	28,357	28,357	28,357	28,357	27,299
Number of id	4,051	4,051	4,051	4,051	3,903
Number of Instruments	14	17	19	19	20
AR2 Test (p-value)	0.965	0.891	0.768	0.424	0.533
Hansen Test (p-value)	0.344	0.274	0.278	0.106	0.112

BIG-DF: La dinamica degli investimenti in Italia

Il modello a correzione dell'errore

Dependent Variable: I_t/K_{t-1}	(1)	(2)	(3)
I_{t-1}/K_{t-2}	0.0568 (0.109)	0.0481 (0.0718)	-0.00223 (0.0682)
Δy_t	0.245*** (0.0262)	0.242*** (0.0259)	0.248*** (0.0264)
Δy_{t-1}	0.292*** (0.0345)	0.297*** (0.0294)	0.310*** (0.0296)
$\Delta TAUC_t$		-0.326*** (0.0182)	
$(k - \gamma)_{t-2}$	-0.289*** (0.0349)	-0.293*** (0.0304)	-0.305*** (0.0306)
Cash Flow _t / K_{t-1}	0.00490 (0.00415)	0.00363 (0.00363)	0.00342 (0.00368)
Cash Flow _{t-1} / K_{t-2}	-0.00417** (0.00194)	-0.00462** (0.00204)	-0.00408** (0.00185)
$\Delta METR_t$			-0.612*** (0.0342)
Observations	96,838	96,838	96,838
Number of id	13,834	13,834	13,834
Number of Instruments	17	19	19
AR3 Test (p-value)	0.682	0.582	0.713
Hansen Test (p-value)	0.667	0.864	0.620