

# Previsioni macroeconomiche e nuovi *leading indicators*

Uso di Big Data per produrre conoscenza orientata  
alle decisioni

Righi A., Iannaccone R., Moauro F.

Forum PA2017

# Indice

1. Scopo dell'intervento
2. Perché i BD possono migliorare le previsioni macroeconomiche
3. Le esperienze internazionali
4. Sperimentazioni Istat di uso combinato di Big Data e dati tradizionali
5. Conclusioni

# Scopo dell'intervento

- Presentare i motivi per cui i BD possono **migliorare** le previsioni macroeconomiche e il nowcasting e aiutare **nell'individuazione** di *leading indicators* per la determinazione dei punti di svolta del ciclo economico
- Fare una panoramica delle **applicazioni realizzate** a livello internazionale
- Presentare alcune **sperimentazioni in corso** in Istat volte al miglioramento del nowcasting del PIL o di altre grandezze macroeconomiche

## Perché i BD possono migliorare le previsioni macroeconomiche

- I contemporanei progressi nell'IT e nella diffusione delle applicazioni Internet consentono al pubblico di accedere ad una grande quantità di informazioni
- I Big data rappresentano ormai un'importante fonte di informazione che può essere utilizzata a fini statistici mediante lo sviluppo degli opportuni metodi
- Considerando la crescente complessità della società e il ritmo in cui la società cambia, essi possono essere molto utili per l'analisi di relazioni complesse e per la produzione di informazione statistica quasi in real-time
- Oggi gli utilizzatori non richiedono solo dati più tempestivi ma “conoscenze” basate su dati di qualità sempre migliore e in grado di orientare le decisioni

# Perché i BD possono migliorare le previsioni macroeconomiche

Anche le previsioni macroeconomiche e l'individuazione di indicatori anticipatori del ciclo economico sembrano beneficiare delle opportunità derivanti dall'utilizzo delle nuove fonti Big Data e/o dall'uso di queste insieme alle fonti più tradizionali

Per Eurostat, il nowcasting di indicatori macroeconomici, in particolare, è il campo in cui i BD possono giocare un ruolo decisivo in futuro (vedi: Big Data and Macroeconomic Nowcasting: from data access to modelling, 2016)

L'uso di BD consente, infatti:

- monitoraggio delle tendenze macroeconomiche (es. dati di Google per la previsione o analisi di *sentiment* dei consumatori)
- monitoraggio della stabilità finanziaria (*sentiment* nei mercati finanziari, incertezza) e non (incertezza)
- calcolo di indicatori di «allerta» precoce

## Perché i BD possono migliorare le previsioni macroeconomiche

- Da quando, nel 2011, Choi e Varian hanno riscontrato che i dati ottenuti da *query* di Google possono essere *leading indicators* di successivi acquisti, perché i consumatori pianificano in anticipo i loro consumi, diversi studi sono stati realizzati in questa area
- Anche l'Istat si muove su questa strada, aprendosi allo sviluppo di nuovi indicatori da affiancare a quelli tradizionali derivanti da indagini, pur continuando ad assicurare la qualità delle statistiche ufficiali prodotte
- Non è infatti sufficiente individuare le possibili nuove fonti, ma bisogna considerare anche le sfide inerenti la preparazione, pulizia, filtraggio e valutazione dei dati

# Le esperienze internazionali

In letteratura esistono due approcci per queste «nuove» previsioni:

1. stime real-time che fanno uso di informazioni provenienti da ogni nuova release di dati
2. nowcasting utilizzando *leading indicators* derivanti da BD (Google trends, Twitter)

L'OCSE definisce un *leading indicator* come :

- una serie temporale che presenta una relazione ragionevolmente coerente con la serie di riferimento per il ciclo macroeconomico in un paese
- un indicatore composito o un insieme di indicatori che possono aiutare a anticipare i punti di svolta nell'attività economica

# Le esperienze internazionali

1. **Stime real-time** che fanno uso di informazioni provenienti da ogni nuova release di dati

Es.:

- Giannone, Reichlin, Small (2008), Nowcasting: The real-time informational content of macroeconomic data, *Journal of Monetary Economics*, 55
  - Higgins (2014), GDPNow: A Model for GDP “Nowcasting” FEDERAL RESERVE BANK of ATLANTA, WORKING PAPER SERIES
  - Carriero, Clark, Marcellino (2014), Real-Time Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility
2. Choi e Varian hanno riscontrato che i dati ottenuti mediante *query* di **Google** possono essere *leading indicators* per :
    - vendite di auto
    - disoccupazione
    - pianificazione di viaggi
    - fiducia dei consumatori

# Le esperienze internazionali

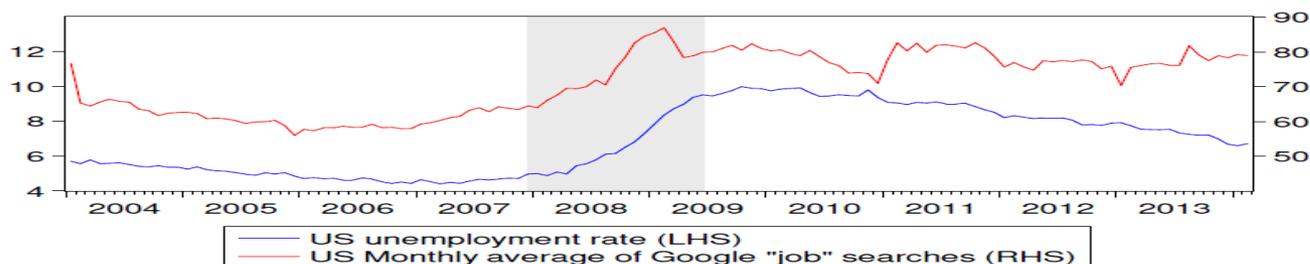
## Filone nowcasting con BD (Google trends)

Perché si utilizza Google? Secondo comscore.com tra 2004 e 2016 Google è dominante nel mercato dei motori di ricerca (56% - 2004 e 64% - 2016)

- Google trends mostra il numero di ricerche effettuate per una determinata parola chiave rispetto al totale delle ricerche  i dati indicano la probabilità relativa che un utente ricerchi una determinata parola chiave in un certo momento
- Sono raccolti:
  - ✓ utilizzando le informazioni di indirizzo IP dai logs di Google e aggiornate quotidianamente
  - ✓ solo se il numero di ricerche supera una determinata soglia di traffico
  - ✓ ma vengono eliminate le *query* ripetute da un singolo utente
- Sono disponibili per paese, regione, città e vengono normalizzati (suddivisi per il traffico totale per area geografica)
- L'indice del volume di ricerca viene scalato dividendo ciascun punto dati dal massimo nella settimana o nella giornata (solo negli ultimi 90 giorni)

## Le esperienze internazionali- Nowcasting con BD (Google trends)

- Koop, Onorante (2013), Macroeconomic Nowcasting Using Google Probabilities, ECB
- Askitas, Zimmermann (2009), Google Econometrics and Unemployment Forecasting. *Applied Economics Quarterly*, 55
- **D'Amuri, Marcucci (2012)** osservano l'associazione tra tasso di disoccupazione USA e ricerche relative alla parola "Job" su Google Trends e suggeriscono l'uso di un indicatore aggiuntivo per prevedere il tasso di disoccupazione mensile, un indice Google (GI) basato sulle ricerche relative al lavoro ottenute da Google trends

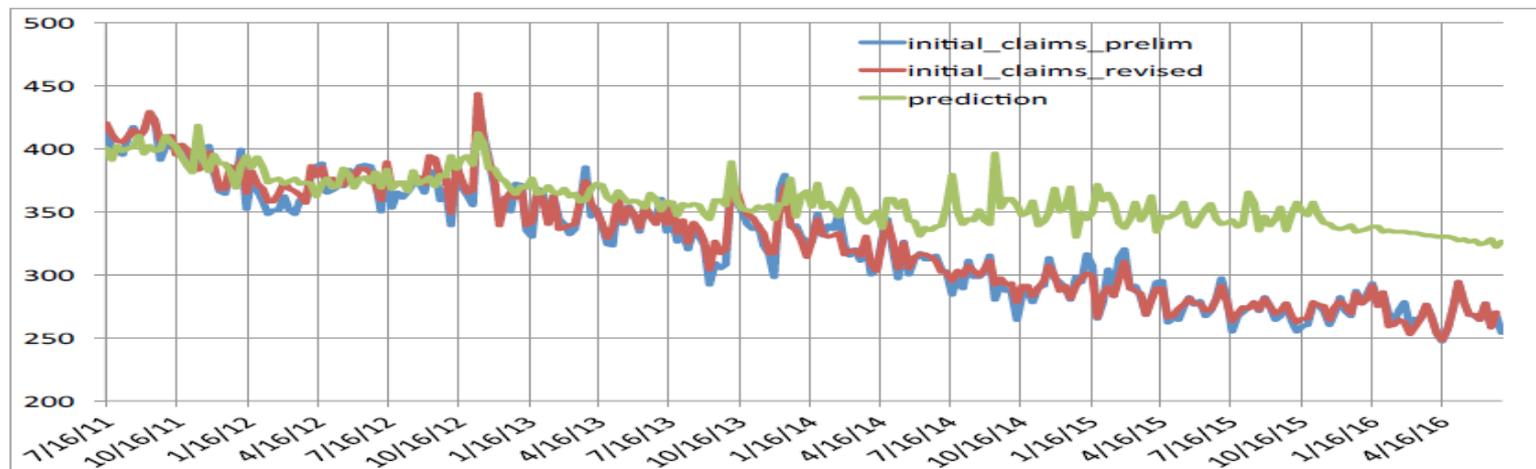


Poi confrontano il potere predittivo dei modelli di previsione lineari utilizzando il *Google index* o modelli che usano variabili più tradizionali (richieste di indennità disoccupazione, Aspettative dei consumatori, Aspettative occupazionali settoriali) e scoprono che il nuovo indicatore migliora le previsioni più degli altri

## Le esperienze internazionali - Nowcasting con BD (Twitter)

Antenucci et al. (2013):

- prevedono la Job loss (espressa in termini di US Initial Claims for Unemployment insurance) con Twitter
- I testi dei messaggi Twitter sono analizzati per individuare la stringa «lost my job»
- Utilizzano i dati per produrre un *real-time Social media index*



# Le esperienze internazionali

<b>Galbraith J.W. and G. Tkacz (2015), Nowcasting GDP with electronic payments data, European Central Bank (ECB). Statistics Paper Series No 10 / 2015</b>	<b>Mostrano l'utilità di un ampio set di dati di pagamenti elettronici (comprese transazioni di carte di credito e debito e assegni) come indicatori del PIL (tasso di crescita trimestrale) Queste variabili catturano un'ampia gamma di spese e sono disponibili in modo molto tempestivo Mentre ogni transazione effettuata con questi meccanismi di pagamento è osservabile, i dati vengono aggregati per la previsione macroeconomica Tra le variabili di pagamento considerate, le transazioni con carta di debito sembrano produrre i maggiori miglioramenti nella precisione previsiva</b>
--	--

- Per una review generale

<b>Hassani H., Sirimal Silva E. (2015) Forecasting with Big Data: A Review, Ann. Data. Sci. (2015) 2(1):5–19 DOI 10.1007/s40745-015-0029-9 © Springer-Verlag Berlin Heidelberg</b>	<b>Si identificando i problemi, le potenzialità, le sfide e le relative applicazioni La rassegna rileva che al momento i settori dell'economia, dell'energia e della dinamica della popolazione sono i principali sfruttatori di BD per le previsioni, e i Factor models, i modelli bayesiani e le reti neurali sono gli strumenti più comuni adottati per la previsione con BD</b>
--	---

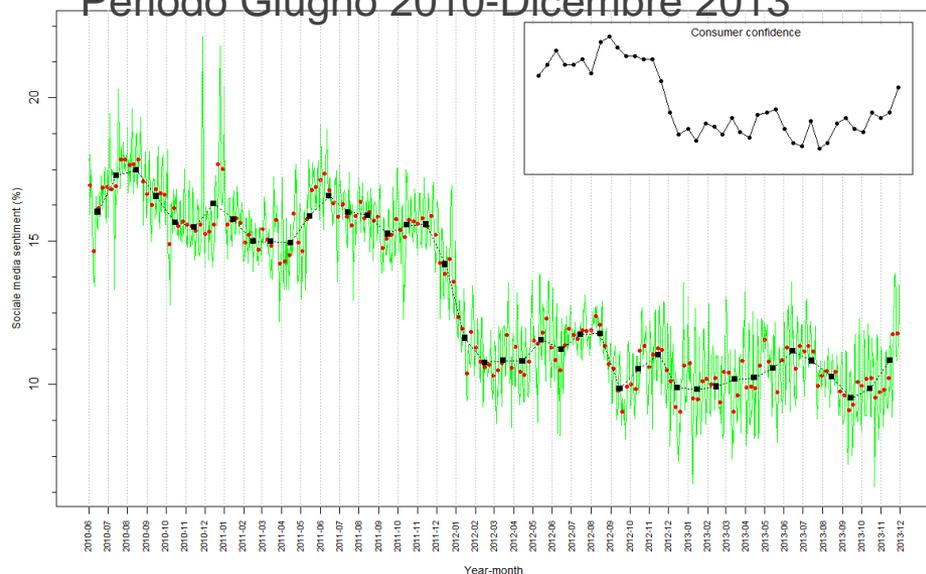
# Sperimentazioni Istat di uso combinato di Big Data e dati tradizionali

Le sperimentazioni in corso per migliorare le stime macroeconomiche:

- studio della Fiducia relativa alla situazione economica (o suoi sotto insiemi) e studio dell'incertezza in base a messaggi Twitter
- uso di dati sul traffico pesante in base a dati dei gestori rete autostradale Aiscat
- uso dei dati dei Pagamenti elettronici e da carte di credito in possesso della Banca d'Italia
- uso dei Consumi elettrici privati da smart meters del gestore del Acquirente unico (in fase esplorativa)
- uso di Google trends per previsione disoccupazione e spese per consumi

# Sperimentazioni Istat di uso combinato di Big Data e dati tradizionali

**Figura** – Serie di fiducia consumatori da Social media (Facebook, Twitter, blogs,...) in Olanda  
Andamento giornaliero (verde), settimanale (rosso) e mensile (nero)  
Periodo Giugno 2010-Dicembre 2013



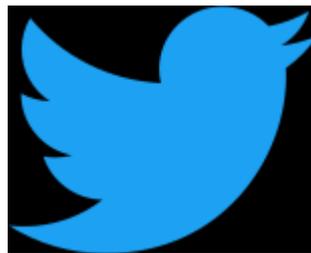
Nell'inserto c'è la serie mensile ufficiale del Consumer confidence CBS

- Lo studio olandese (CBS) rivela:
- ❖ una chiara **associazione** tra i cambiamenti nel Sentiment calcolato dai messaggi dei Social media e il Clima di fiducia dei consumatori ufficiale
  - ❖ I cambiamenti nel Sentiment dei Social media precedono sempre quelli nella Consumer confidence con un lag di 7 giorni
  - ❖ I messaggi che influenzano di più l'indice dei Social media sono quelli Facebook e poi quelli Twitter

# Sperimentazioni Istat di uso combinato di Big Data e dati tradizionali

Obiettivo: miglioramento delle stime anticipate del tasso di crescita del PIL con sentiment di fiducia nella situazione economica o, al contrario, di incertezza

- 1) Produzione di serie di indicatore di *sentiment* da Social media – Twitter a cadenza inferiore al mese
- 2) Confronto con le serie del clima di fiducia consumatori
- 3) Arricchimento delle serie del modello economico



# Sperimentazioni di uso combinato di Big Data e dati tradizionali

Obiettivo : miglioramento delle stime anticipate del tasso di crescita di alcune componenti del PIL per mezzo dei dati sul **traffico pesante autostradale**

Stime di veicoli / Km per tipo veicolo sulle principali autostrade italiane, dati trimestrali per 60 tronchi autostradali (dati Aiscat) dal 1995 (ove possibile dati mensili per società concessionaria)

Fase 1 – confronto delle serie trimestrali (sulla base di dati non Big) con serie del GDP a fini di nowcasting o previsione

# Sperimentazioni di uso combinato di Big Data e dati tradizionali

## Analisi di correlazione

Serie: Aiscat e Contabilità Nazionale a prezzi concatenati (2004:4- 2016:2)

	Traffico Pesante	Totale traffico (leggero+pesante)
<b>PIL</b>	<b>0.387</b>	
<b>Totale Valore aggiunto</b>	<b>0.380</b>	
<b>Valore aggiunto branca commercio, trasporti, alloggi e ristorazione</b>	<b>0.360</b>	
<b>Valore aggiunto Servizi</b>		<b>0.620</b>

In fase di studio l'analisi di impatto delle serie di alcune tratte specifiche sul nowcast delle componenti dei servizi e del PIL:

- Le tratte più correlate al PIL sono **Ivrea-Torino e Ivrea-Santhià (0.8)**
- seguite da Firenze-Pisa, Nord Roma-Civitavecchia, Roma-Pescara, Nodo di Mestre, Quincinetto-Aosta e Messina-Catania

# Sperimentazioni di uso combinato di Big Data e dati tradizionali



## Information (Metadata)

**Road: A24**  
**Km: 6,9**  
**Date: 31/01/2014**  
**Time: 11:42:13**

## Value (Data)

**Vehicles: 8**

Fase 2 – Utilizzo di serie di BD sui flussi di traffico con dati da immagini (in corso di sperimentazione) a fini di confronto o di nowcasting

L'utilizzo di immagini da telecamere con riconoscimento automatico dei veicoli permetterà di:

- stimare i flussi di traffico stradale/autostradale di tutti i veicoli
- di avere indicazioni su vasta scala in tempo reale disponibili anche per le previsioni macroeconomiche

# Sperimentazioni di uso combinato di Big Data e dati tradizionali

Obiettivo : miglioramento delle stime anticipate sia del tasso di crescita del PIL attraverso dati dei flussi dei **pagamenti elettronici e di carte di credito** del sistema dei pagamenti della BCE

Fase 1 – utilizzo delle serie aggregate relative ai Circuiti di scambio e regolamento interbancario del Sistema dei pagamenti internazionale su transazioni elettroniche a fine di previsione delle serie del GDP, o di altri macro aggregati

Fase 2 – in collaborazione con la Banca d'Italia, si vorrebbe produrre delle stime delle serie delle transazioni elettroniche per la stima delle vendite totali o per tipologie (e-commerce, spese turistiche, etc.) o per la previsione delle componenti del PIL o dei consumi

# Sperimentazioni di uso combinato di Big Data e dati tradizionali

## Fase 1 - Analisi di correlazione

Serie differenziate: Banca d'Italia e Contabilità Nazionale (2005:1- 2016:3)

	<b>Circuiti di scambio e regolamento interbancario</b>	
	<b>Sistema BI-COMP Flussi lordi (BI-REL e TARGET2-Bdl) (recapiti locale)</b>	<b>Regolamento lordo di cui Clientela</b>
<b>PIL</b>	<b>0.791</b>	<b>0.759</b>
<b>Totale Valore Aggiunto</b>	<b>0.772</b>	<b>0.787</b>
<b>Valore Aggiunto Commercio</b>	<b>0.644</b>	<b>0.641</b>
<b>Valore aggiunto Servizi</b>	<b>0.653</b>	<b>0.757</b>

# Sperimentazioni di uso combinato di Big Data e dati tradizionali



Obiettivo : miglioramento delle stime anticipate dell'occupazione per mezzo di indicatori di Google trends

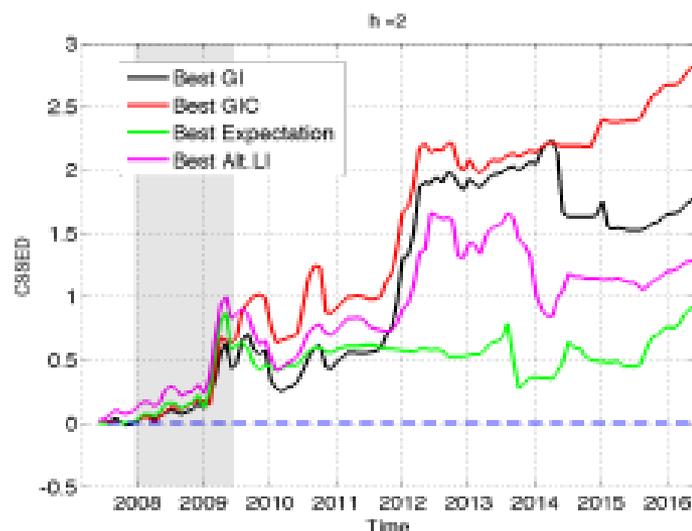
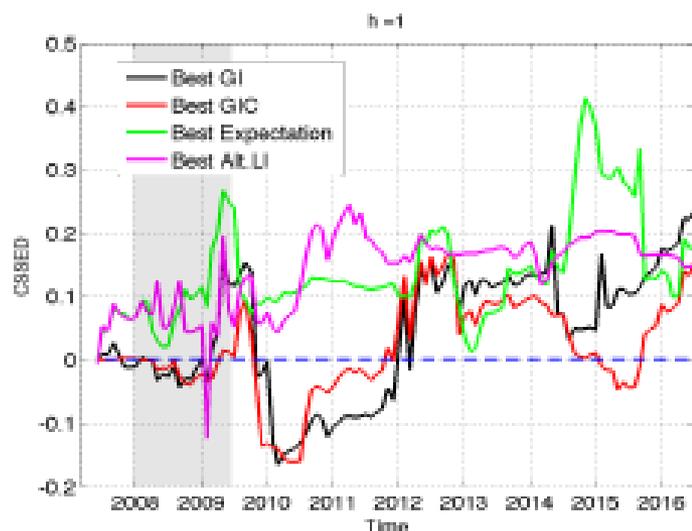
*Precedenti: Askitas et al.; Choi and Varian; D'Amuri and Marcucci; Ginsberg et al.*

*Sfrutta la capacità predittiva dei dati di ricerca web di Google Trends (GT) per l'aggiornamento di rilevanti indicatori economici congiunturali relativi al lavoro*

Sperimentazione Istat-Banca d'Italia in corso produce un Indice di ricerca di lavoro (GI) basato sulle ricerche di offerta di lavoro. Alla serie delle ricerche su «cerco lavoro» si sono aggiunte quelle relative ad altre parole chiave significative o aggiungendo il nome dei principali siti di ricerca di lavoro (subito.it lavoro, indeed, infojobs, jobrapido, kijiji lavoro)

19 Gli indici che usano più termini si dimostrano i migliori in termini previsivi

# Sperimentazioni di uso combinato di Big Data e dati tradizionali



I risultati di Marcucci e Fasulo per la situazione italiana mostrano che nei modelli di nowcasting (previsione a un mese) l'introduzione di serie Google (GI) non migliora la performance predittiva, ma a partire da previsione a 2 mesi invece il modello con GI offre sempre una migliore performance previsione

# Sperimentazioni di uso combinato di Big Data e dati tradizionali



Obiettivo: il miglioramento delle stime anticipate delle spese per consumi attraverso Google trends

In letteratura diversi esempi di nowcasting/forecasting di alcuni indicatori congiunturali (consumi, vendite, viaggi....)

- Vosen, Schmidt (2011), Forecasting private consumption: survey-based indicators vs. Google trends, *Journal of Forecasting*
- Bortoli, Combes (2015), Contribution from Google Trends for forecasting the short-term economic outlook in France: limited avenues, *Conjoncture in France*, INSEE

La sperimentazione di Fasulo, Guandalini, Terribili (presentata a SIEDS, 2017) realizza il nowcasting delle spese trimestrali delle famiglie sfruttando le serie di Google trends per alcune parole chiave (riferite ai termini più ricercati per i principali capitoli di spese considerati dall'indagine Istat)

# Sperimentazioni di uso combinato di Big Data e dati tradizionali

## Serie storica Consumi:

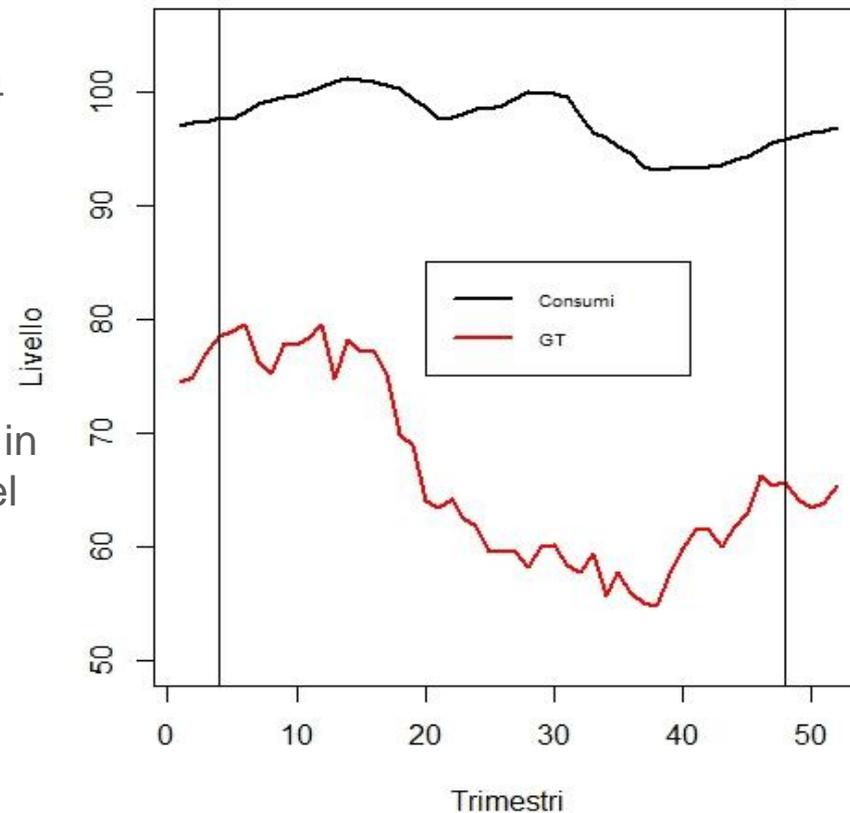
Serie trimestrale destagionalizzata dei consumi delle famiglie Italiane dal 2004

## Serie storica GT:

Serie destagionalizzata (X-13ARIMA-SEATS)

- ✓ Si riesce a prevedere il trimestre in corso utilizzando dati di spesa del trimestre precedente e le serie Google riferite al primo mese del trimestre
- ✓ Il modello con Google Trends ha prestazioni migliori rispetto ai modelli predittivi che sfruttano *leading indicators* classici (fiducia)

**Serie Consumi vs GT**



# Conclusioni

- Big Data sono un'ottima fonte per proxy per aspettative e sentiment che possono anche essere generate ex post e costituiscono la base per nuovi *leading indicators* utili e anche a fini previsivi
- Provenendo da fonti non convenzionali, forniscono una grande quantità di dati (a livello micro) che possono essere un ottimo complemento alle statistiche ufficiali
- Ancora ulteriori opportunità vengono dall'Internet of Things
- Sfruttare queste nuove fonti di dati tempestive possono portare beneficio sia alle analisi sia alle policy