Stefano Nardelli
European Central Bank
Monetary Policy Strategy Division

discussion of the paper

The Potential of Big Housing Data

by M. Loberto, A. Luciani, M. Pangallo

Harnessing Big Data & Machine Learning Technologies for Central Banks
Rome, Banca d'Italia, 26-27 March 2018

RECAP OF RESULTS

- Source: a dataset of **more than 1,000,000** online sales ads for residential units (immobiliare.it)
- Sample: Jan 2015 to Jun 2017 (virtually “real time”)
- Coverage: all Italian provincial capitals.
- Advantages:
  - ✓ overcome the limitations of existing administrative data
  - ✓ physical characteristics of housing units
  - ✓ proxy of the **time on market** (i.e. #days the housing unit on sale as listed on the website) as well as proxy for the **demand tightness** at the level of individual neighborhoods (i.e. average #visits (clicks) on ads in specific areas)

RECAP OF RESULTS

- Validation against **official statistical sources** (i.e. Osservatorio del Mercato Immobiliare database, but low frequency, non public and limited info on physical characteristics; and quartely Italian Housing Market Survey by BdI, OMI and Technoborsa)
- core indicators on Italian housing market from OMI vs imm.it match
- Sampling issue (on *imm.it*): a **sizeable number of duplicates** exist, since no limitations on ads for the same housing unit/sellers
- Duplicates identified with **machine learning techniques**
- Final “clean” dataset of housing units contains only **~63% of effective dwellings** out of total ads
(**high heterogeneity** by area/“asset liquidity”)

Potential applications:

- ✓ **nowcasting** of aggregate / local price trends
- ✓ the detailed study of **heterogeneity** and **market segmentation**
- ✓ **quality-adjusted price index** (for Rome and Milan but applicable also at national level) with hedonic regression to account for physical characteristics of housing units
- ✓ Preliminary evidence that **#visits (i.e. clicks) on ads** for dwellings located in a specific neighbourhood **leads prices** in the same area
- ✓ **Demand tightness** has predictive power on price trends at the national / local level thus may inform policies dealing with the construction and financial industries

Is the study **relevant**?

- ✓ For **economic theory**, "*one major problem is that dwellings are heterogeneous goods exchanged in decentralized and segmented markets. […] spatial and informational frictions make the Walrasian equilibrium concept unsuitable for this market*"
- ✓ also for **empirical application**, "… *microdata on actual housing transactions are available only in a few countries and unfortunately many of these sources show limitations in the spatial or in the temporal dimension …*"
- ✓ as well as "… *much more challenging is finding comprehensive information about the full history of housing transactions, from the moment the dwelling goes on the market up the actual transaction*"

Is the study **relevant**?

- ✓ Also for **policy actions**, in particular from a **macroprudential perspective** as it is key to promptly identify and properly address the emergence of real estate bubbles
- ✓ but also for **monetary policy** (especially when other policies fail)
- ✓ At European level, the ESRB recommended to close real estate data gaps and the ECB Statistics Committee established a real estate task force (RETF) to explore feasibility of sourcing a range of **financial real estate indicators** to fill in the identified data gaps and better **support future more evidence-based policy discussions on real estate developments**
- ✓ Focus on banks but key to have hard data on real estate market

RELEVANCE

Information area	Market participants	Countries for which information on the indicator is available / largely available	
		RRE	CRE
Lending - flows **Area:** assessing the respective dynamics of financing supply and demand.	**Banks**	**AT, BE, CY, CZ, DE, DK, EE, ES, FR, GR, IE, IT, LT, LU, LV, MT, NL, PT, PL, SE, SK**	**AT, CZ, EE, ES, FR, NL, PT**
	Insurance companies and pension funds	**BE, GR, NL**	
	Other market participants	**BE, NL**	**NL**
Lending - stocks **Area:** assessing the respective levels of financing supply and demand	**Banks**	**AT, BE, CY, CZ, DE, DK, EE, ES, FR, GR, IT, LT, LU, LV, MT, NL, PT, PL, SE, SI**	**AT, BE, CZ, DE, DK, EE, ES, FR, GR, IT, LT, LV, MT, NL, PT, PL, SE, SI**
	Insurance companies and pension funds	**AT, DE, GR, LV, MT, NL**	**AT, BE, DE, FR, LV**
	Other market participants	**LV, NL, PT**	**LV, NL, PT**
Lending - NPLs and impairments **Area:** monitoring the credit quality	**Banks**	**BE, CZ, DE, EE, ES, FR, GR, IE, IT, LT, LU, LV, MT, NL, PT, PL, SE, SI, SK**	**BE, DE, EE, ES, FR, GR, IT, LT, LV, MT, NL, PT, PL, SI**
	Insurance companies and pension funds	**LV**	**LV**
	Other market participants	**LV**	**LV**
Lending - lending criteria **Area:** tracking the evolution of lending conditions	**Banks**	**AT, BE, CY, CZ, EE, ES, FR, IE, IT, LT, MT, NL, PT, PL, SE, SK**	**AT, CZ, DK, ES, FR, IE, MT, NL, SE**
	Insurance companies and pension funds	**AT**	**AT**
	Other market participants		

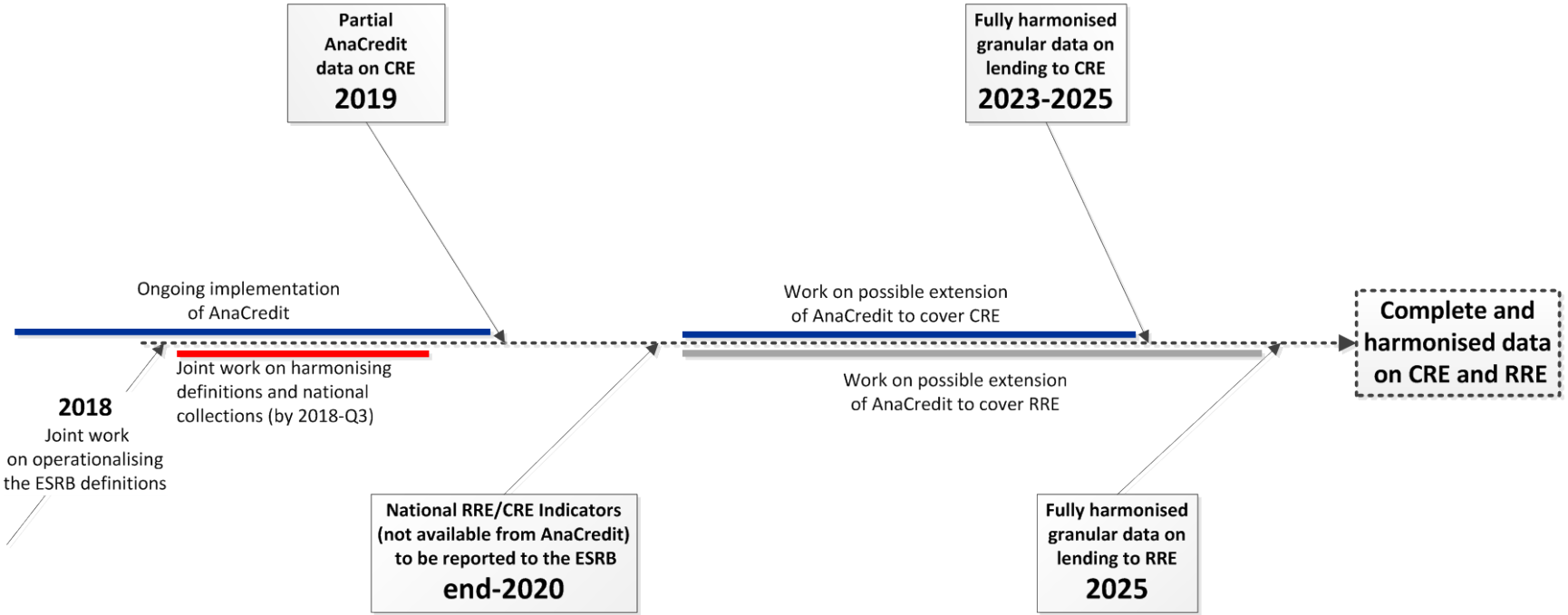
Availability of financial system's data on residential and commercial real estate by information areas

Note: The availability of the data considers whether the data includes enough details to prepare the most relevant indicators in the given information area. For example, available/largely available data stands for a situation where all or the most relevant breakdowns/indicators can be compiled from existing sources.

Availability of financial system's data on investments in real estate by information areas

Information area	Market participants	Countries for which information on the indicator is available
Investments - stocks **Area:** assessing the dynamics of CRE investments	**Banks**	**DE, DK, EE, ES**
	Insurance companies and pension funds	**AT, BE, DE, EE, FR, GR, NL, SE**
	Investment funds	**AT, DK, EE, GR, IE, NL**
	Other market participants	**BE, EE**
Investments - flows **Areas:** assessing the levels of CRE investments	**Banks**	
	Insurance companies and pension funds	**GR, NL, SE**
	Investment funds	**GR, IE, NL**
	Other market participants	

RELEVANCE


Partial AnaCredit data on CRE
2019
Fully harmonised granular data on lending to CRE
2023-2025
Ongoing implementation of AnaCredit
Joint work on harmonising definitions and national collections (by 2018-Q3)
2018
Joint work on operationalising the ESRB definitions
Work on possible extension of AnaCredit to cover CRE
Work on possible extension of AnaCredit to cover RRE
Complete and harmonised data on CRE and RRE
National RRE/CRE Indicators (not available from AnaCredit) to be reported to the ESRB
end-2020
Fully harmonised granular data on lending to RRE
2025

DATA ISSUES

- Real estate market indicators should be:
  1. **Accurate**
  2. **Timely**
  3. **Representative**
  4. **Methodologically consistent** (at area-wide level)

- Methodology presented offers several advantages especially for **2.** (from biannual/quartely to real time as well as significant reduction of publication delays) and **3.** (deeper granularity)
- Virtually for **4.** if applied to other countries (e.g. immobilienscout24.de, www.immobilier-france.fr, etc.)
- While **1.** difficult to assess for the lack of reliable sources

DATA ISSUES

Issue: actual transaction prices unavailable (also in official sources)

- Estimate of **discounts** usually between 5-10% of ask price (Italian Housing Market Survey, Nomisma) but
  - ✓ not linked to typology of dwelling?
  - ✓ independent of business cycle/market liquidity?
  - ✓ reverse engineering conceivable to derive price from indirect estate market depth measurement (i.e. bid-ask spread)

- Benchmarking against broader **macroeconomic factors** may offer a better founded validation of results (e.g. factors driving demand such as income or unemployment; financial conditions and regulations; monetary policy cycle)

General comment

- **Interesting** and **promising approach**
- but **streamlining** and **focusing on methodology** rather than possible use would sharpen the message(s)
- Very precise on technical descriptions but not too compelling on potential applications
- In particular, Section 5 offers (too) many good ideas on possible developments but not sufficiently developed (e.g. atypicality)