Granular data and macroprudential policy: examples and challenges

Klaas Knot

Introduction

Granular data have become more relevant in recent years. One momentous event in which granular data played an important role was the comprehensive assessment in the run-up to the SSM, and more specifically the asset quality review (AQR) that was part of it. In what still seems like an impossibly short period of time, banks and NCBs had to learn how to supply, receive, process, validate and analyze huge amounts of loan-by-loan data. This highly demanding exercise heralded a new style of supervision on financial institutions, one that would be more data-driven than we were used to. Another example is a practice that was introduced at De Nederlandsche Bank several years ago, when we started requesting, at regular intervals, granular datasets with loans from financial institutions under our supervision in order to validate their risk models.

With the collection of Money Market Statistics under regulation ECB/2014/48, starting 1 April 2016, and the future collection of and Analytical Credit datasets (Anacredit), granular data will gain further in importance. Not just in banking supervision, but also in monetary policy and macroprudential policy. This marks no less than a radical change in collecting, analyzing, managing and providing data. One that will alter the prevailing views, mindsets and modes of operation of statisticians and users. In view of this impact, this article will look into the benefits and value added of granular data for macroprudential policy, partly on the basis of concrete examples of how De Nederlandsche Bank uses microdata in performing its financial stability task. Subsequently, it will reflect on the challenges posed by such datasets. It will conclude by presenting some caveats.

The benefits of granular data

Granular data make for a more accurate assessment of systemic risks as well as a more precise calibration of macroprudential instruments.

The relevance of microdata for assessing systemic risks is best illustrated by the following two examples.

The first example relates to the risks of Dutch mortgages. The table below gives some aggregate indicators (for end 2013) as are typically used to assess risks or imbalances:

Household mortgage debt to disposable income	108%
Loan to deposit ratio of the domestic banking sector	175%
Average LTV of first-time buyers	95% (105% at peak)
Share of interest-only mortgages	58%
Credit losses	12 bps

Alarming as the aggregate mortgage debt characteristics may look, credit losses have remained quite low even after a severe housing market bust and a nominal price decline of more than 20%. Evidently, on the basis of these macrodata, it is hard to draw any firm conclusions regarding systemic risk, let alone the suitable macroprudential policies. To better understand the risk characteristics of the mortgage debt, we need to dig deeper by using micro-data.

In 2012, the ECB launched the *Loan Level Initiative*, under which banks wishing to collateralize RMBSs with the ECB were required to provide granular data on the underlying mortgages. As they had to supply this information to the European Data Warehouse, we requested these banks to also report these data to DNB, on a quarterly basis. Moreover, we asked them to supplement the data on securitized loans with granular data on non-securitized mortgages. It turned out that the granular dataset thus collected proved quite useful, as will be illustrated below.

We derived the following valuable insights from these loan-level data.[1]

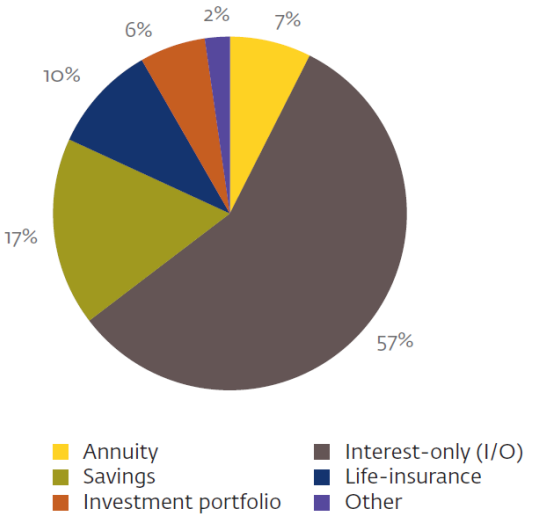
- On interest-only mortgages: 60% of the Dutch mortgage portfolio consist of interest-only loans (Figure 1, left-hand panel). The loan level data enabled us to see how these interest-only loans are distributed across the population. We found that about 25% of the portfolio consist of fully interest-only mortgages, and that over 60% of the households combine an interest-only

[1] See for more details: Mastrogiacomo and Van der Molen (2015), Dutch mortgages in the DNB loan level data, DNB Occasional Study 13-4.

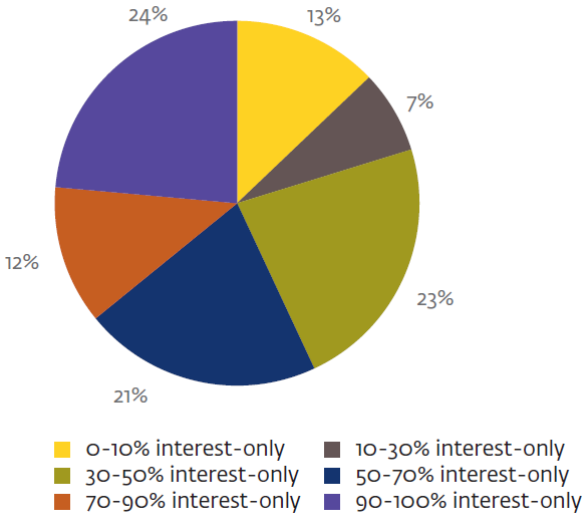
loan with an amortizing loan (Figure 1, right-hand panel). This is quite relevant information from a risk perspective.

Figure 1: interest-only loans

By loan type

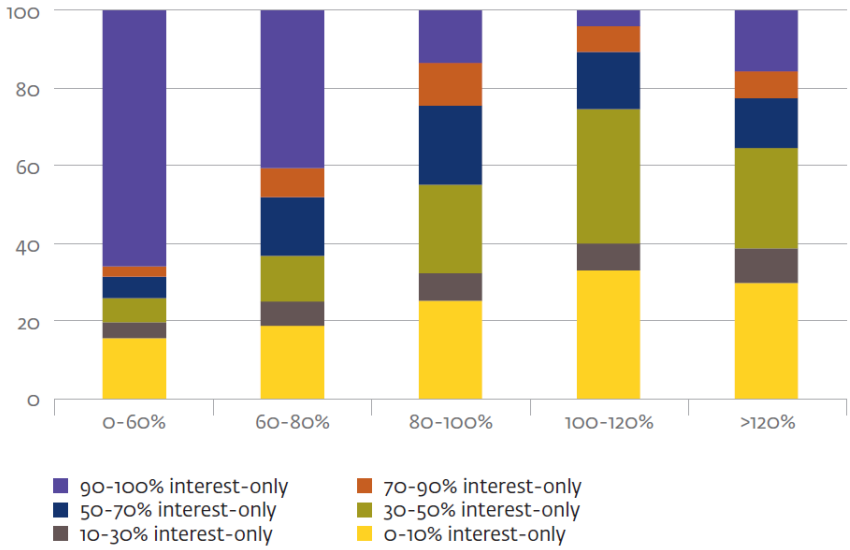


By interest-only component



- On the correlation between different risk characteristics: A high share of interest-only loans and high loan-to-value ratios looks like an especially risky combination. The loan-level data showed us that in fact the two risk characteristics are negatively correlated. Fully interest-only mortgages typically have low LTV-ratios, whereas high-LTV mortgages often have a relatively low interest-only component (see Figure 2).

Figure 2: LTV buckets and interest-only component



Note: LTV-buckets are on the x-axis, columns show the distribution of loans across different interest-only components.

A second example of the usefulness of microdata for assessing systemic risks is macro-stress testing.

In a stress test, we seek to determine how losses of, e.g., a bank are affected in an adverse economic scenario. One way of doing so is to use historical data and regression methods designed to estimate directly the sensitivity of the risk parameters (like probability of default [PD] and loss given default [LGD]) to certain macroeconomic variables (GDP growth, unemployment rate or house prices).

The availability of individual household balance sheet data enables us to follow an alternative approach. To model the impact of a stress scenario, we simulate how households' income, debt service ratios and mortgage characteristics would be affected by, e.g., unemployment and a decline in house prices. From these types of links and using micro-econometric simulation techniques we deduce how PD and LGD would be affected in this scenario and what magnitude of loan losses may be expected.

Working from the household balance sheet has two distinct advantages. First, we use entirely different data sources and techniques than a bank would

typically use in a bottom-up stress test approach. Using these data and methods provides an independent benchmark for the risk modelling done by banks.

Second, and more importantly, this approach is much better suited to the spirit of scenario analysis. Using historical data on bank loans gives answers to questions that are backward-looking by nature, e.g., “What is the risk that this loan defaults, given the historical default probability of similar loans under stress?" By contrast, our approach permits us to investigate what happens when assumptions taken for granted are dropped and replaced by other heuristics, like “What if households that over the next ten years cannot hope to recover from their negative home equity, start to default on their loans, even though they have kept up with their payments in the past?" In other words, simulations encourage us to think about truly extreme scenarios, and enable us to assess the impact of tail events: events with a low probability but a large impact.

But microdata are not only crucial for accurate systemic risk assessment. We also need microdata for assessing the impact of potential macroprudential policies.

This is best illustrated with a recent analysis from DNB of the effects of a lower LTV limit. The LTV-limit allowed in the Netherlands is high. Currently, the legal maximum LTV of a new mortgage is 102%, and will be 100% in 2018. In DNB’s view, that is still too high, and a further reduction is needed. But: to what level and at what pace?

Microdata again are crucial in answering these questions. First, a reduction in the LTV-limit requires buyers to make a larger down payment. In the short term, this will have a negative impact on transactions and, consequently, house prices, increasing the demand for rental housing. To estimate the magnitude of these effects, we combined data on original LTVs of mortgages and on individual households financial assets in order to determine which households would be restricted by a reduction in the LTV limit. Moreover, we used income and balance sheet data to estimate the time it would take these households to accumulate enough savings to be able to enter the housing market. On the basis of this estimation we concluded that a gradual reduction of the LTV-limit to 90% would

enable an orderly transmission of the housing market to a more stable structural situation.

We also used microdata to analyze the potential benefits of a lower LTV-limit. For instance, a lower LTV-limit reduces the probability of negative home equity, or, underwater mortgages. Using the loan-level data, we were able to make simulations to establish the relation between the level of the LTV-limit and the probability of being underwater after a severe housing market correction. This analysis showed that a reduction of the LTV-limit from 100% to 90% substantially reduced the probability of being underwater.[2]

Clearly, granular data – if set up properly – may serve a great many different purposes and, hence, in the long run yield cost savings for financial institutions. These were the reasons behind the start of the AnaCredit project. Potentially, AnaCredit is suitable for a variety of purposes, i.e. statistics, monetary policy and its implementation (incl. collateral management), financial stability, research and - in the longer term - also supervision. The use of granular data compels NCBs to become more flexible in addressing topical issues, and largely abandon the use of ad hoc datasets.

The measure of detail contained by granular data also enables us to link the collected data to other public and non-public sources. DNB's recent initiative to enhance granular data on mortgages with data available from Statistics Netherlands – the national statistical institute in the Netherlands - serves as a clear example. Statistics Netherlands possesses a wealth of up-to-date information on Dutch households' income, debt and assets, made available by the Dutch tax authorities. Anonymized, this information may be linked to the loan level data, thus producing a more complete and nuanced picture of Dutch households' debt, income and wealth. This possibility permits new analyses, which may provide new policy insights.

Challenges surrounding granular data

For all their benefits, granular data and the new possibilities they open up for analysis come with serious challenges though. For one, all stakeholders involved

[2] See for more details: Verbruggen et al. (2015), Effects of further reductions in the LTV limit, DNB Occasional Study 13-2.

will need to adopt a new mindset when working with granular data. From the very beginning, statistical reports have been set up for the purpose of compiling aggregates and macroeconomic statistics. The detail and coverage required to this end is limited. With the shift to granular data this will change significantly. Collecting 'data molecules' enables statisticians as well as ordinary users to deduce whatever information they are looking for. For example, detailed information on banks' loan portfolios will allow some users to compile aggregates (i.e. loans to non-financial corporations) and others to analyze the exposure of banks towards loans to carbon intensive industries by geographical location. To be able to use granular data for various end products -be they old-fashioned aggregates or risk metrics of loan portfolios- statisticians will need to describe and classify every single 'molecule' in detail.

Another challenge is that the use and related intercomparability of 'data molecules' between financial institutions, but also between NCBs, call for harmonized data modelling. This means that central banks within the European System will need to develop a single model in collaboration with the reporting institutions, so as to ensure that each and every one observes one and the same definition of a loan and of all the data attributes involved. In addition to furthering the applicability of the data, harmonization will also hugely diminish the costs involved for financial institutions, especially those active in different euro area countries. While the first step towards harmonization is currently being undertaken as part of the Banks' Integrated Reporting Dictionary (BIRD) project, we should not expect full harmonization to be achieved any time soon. It will be a time-consuming exertion, mostly owing to existing country-specific classifications of instruments, entities and attributes. In addition, harmonization on a European scale can also impact the autonomy the individual member states involved. Fostering willingness to give up this freedom for the greater good of more harmonization may well prove
a ponderous process.

A third challenge lies in the circumstance that assessing and processing vast quantities of data and classifying granular data down to the minutest detail require different capacities in terms of automation and human resources than are traditionally available at the central banks and financial institutions. This is

why the industry and NCBs will need to invest heavily in hardware facilitating the collection, storage, management, disseminations and use of data. In anticipation of these requirements, DNB has recently launched a program to design a production system that accommodates these developments. Currently, DNB is tailoring its staff's capacities accordingly. Experts are hired externally and staff members are being trained to develop and acquire the skills and knowledge needed, e.g., for data modeling and (advanced) programming.

Finally, the users, too, will be confronted with some significant changes. The practically limitless options require that they formulate their wishes more critically, asking themselves "what do I really need?"; "what will be my focus?"; "what questions should the data minimally address?". The vast quantity of available data may easily cause users to lose sight of the data they are looking for. In that sense the old maxim "less is more" seems to be gradually replaced by "more is more". But is the latter really what we want? Put differently: both users and statisticians should be critical of using granular data without reserve.

Some caveats

While welcoming the wide array of possibilities inherent in granular data, we should bear in mind that with great power, comes great responsibility. Not just responsibility towards the benefits and costs of granular data for NCBs and financial institutions, but also towards the more ethical and security-related aspects of collecting potentially sensitive information.

Working with granular data is a relatively new discipline. A discipline that, for the present, mainly focuses on the seemingly boundless opportunities offered by granular data. Therein lies its appeal. However, this may easily lead users and statisticians to ignore the ethical and security-related implications of collecting and combining granular datasets, especially when information on private individuals is concerned. Protecting the privacy of citizens should in democracies like ours be a priority. We should critically ask ourselves as central bankers and supervisors to what degree we need information on private individuals, on both anonymized and non-anonymized levels. And if we collect personal data, we should make sure that our data security and data governance meet the strictest standards protecting the privacy of the data subjects.

Speaking of power, another issue comes to mind as well. When collecting and combining granular datasets, as central bank and as supervisor we could end up in the position where we know more or could have known more about the loan portfolios of the banks and debtors involved than the financial institution itself. This abundance of potential information on the institution could result in peculiar situations. For example, in times of crisis or in the event of a financial institution's failure, an inquiry – e.g. by a parliamentary committee – may lead to the conclusion that a supervisor could have known about the problems if he had interpreted or used the data properly. Questions could arise like "Why did the supervisor fail to act on the information which was inside its walls all the time"? Greater powers come with greater responsibilities, which could also make ourselves more vulnerable to criticisms of overextension.

Conclusion

Serving many useful purposes, granular data most certainly merit a prominent role in many areas. However, it takes a joint approach to make sure that the benefits (optimum high-grade datasets) and costs (to the ESCB and the industry) are, and stay, properly balanced and that questions surrounding the confidentiality of data are adequately addressed.