

Reconstructing Inter-industry transactions matrices using transactions data

Johannes T Lumma (The Alan Turing Institute), Vasco Carvalho (University of Cambridge), François Lafond (University of Oxford), Kerstin Hoette (The Alan Turing Institute), Victor Meirinhos (Office for National Statistics (UK))

Recently, the UK Office for National Statistics (ONS) has been working with the UK payment system operator to acquire anonymised and aggregated data focusing on financial transactions between SIC-5 industries.

Starting with an aggregated prototype dataset comprising more than >500m financial transactions worth >£7.5tn between SIC-2 industries ranging from 2015-2022, we assess the feasibility of developing statistics consistent with national accounting standards, but based on population-scale naturally occurring data in real time. In particular, we discuss measurement issues that have to be addressed when comparing to National Accounts.

We provide an initial assessment, a roadmap to tackle key bottlenecks, applications to nowcasting, and a research agenda for the use of this dataset in downstream case studies, such as the impact of Brexit and Covid-19.

From transactions data to national accounts

We start from an initial aggregate dataset produced by the ONS, comprising monthly transactions volumes and values between 1020 industry-region pairs (85 SIC 2-digit industries times 14 NUTS-1 regions or islands), for every month since January 2015.

We start by describing the nature of transaction data derived from the payment system infrastructure, and then identify a number of specific issues (in addition to sampling biases) that make it challenging to retrieve national accounts-like quantities from transactions between bank accounts. A first issue is that input-output tables need to classify each B2B transactions into intermediate consumption or investment, while our data simply shows money flows between industries. Second, in national accounts the output of some industries (retail and wholesale trade in particular) is only their margin, that is, it does not include the value of the goods bought for resale. Input-output tables show the flow of goods between industries even if this flow has taken place through trade intermediaries. By contrast, the payments data available to us shows payments from SIC-aggregated industries to the trade sector, and from the trade sector to supplying industries.

We discuss this issue and outline methods of data reconstruction to align payments data with national accounting principles. Third and relatedly, payments may include credit flows, and more generally the raw payments are far from the accounting conventions used for computing the value added by the financial sector. Fourth, the transaction data include activities from the public and non-market sector, which again obey non-standard accounting conventions. We outline strategies

to align naturally occurring data with national accounts data, in the spirit of Buda et al (2022, Cambridge WP in Economics 2244), but focusing on inter-industry tables.

A nowcasting application

A great advantage of transaction data is that it is very timely. Thus, we can attempt to nowcast key economic statistics that are usually released with substantial delays, such as input-output tables, but also sub-national statistics (industry GVA specifically). To give an example, we will evaluate the prediction error using the years 2015-2018, where official data is available, and nowcast the national inter-industry transaction matrices for 2019-2021. We compare the nowcasting ability of two variables: the total value of transactions, and the count of transactions, showing that transactions count sometimes outperform transactions values by their predictive power.

Exploiting the granularity of transactions data

A second great advantage of transactions data is that they are more fine-grained, both in the time series and the cross-sectional dimension. After we have validated the transactions data-based national and annual tables, we analyse regional tables and monthly fluctuations.

First, we benchmark our transactions-based regional table against the annual regional input-output tables available for Scotland for 2015-2019 and regional macroeconomic accounts. Existing regional tables often rely on the imputation derived from national statistics or regional surveys. We assess the ability of our data to produce regional statistics.

Second, monthly data on pairwise transactions between sectors is rare. How volatile are these values? We provide a description of the relationships between key moments of the data and compare these to what has been reported in other datasets. Classic examples include the relationships between the mean and standard deviation of growth rates, or the standard deviation of growth rates and sector sizes.

Towards substantive economic applications

Finally, we discuss more downstream applications. First, how useful would have been transactions-based matrices during Covid lockdowns? A number of papers have estimated the initial shocks to specific industries, and used input-output tables to propagate these initial shocks and predict overall effects. Would these models have worked better with the payments-based input-output tables?

Second, what can be said about the effect of Brexit on supply chains? While our data covers only national payments, we can check whether tradable sectors have had higher levels of rewiring than non-tradable sectors, and whether non-tradable sectors that are highly dependent on tradable sectors have been more affected than those which are not.

Summary

Based on a prototype dataset, we translate financial transaction data into national accounting-like quantities and discuss applications to the development of finer-grained statistics, nowcasting, and economic modelling.