Non-survey regionalization with commodity balance and the gravity model

Topic: Regional Input-Output Economics - III
Author: Mattia Cai

Very few National Statistical Offices produce survey-based input-output (IO) data with subnational (â€˜regionalâ€™) resolution. Regional analysts have debated ways of getting around this problem for several decades now, giving rise to a sizable and constantly expanding literature on IO regionalization.

The greatest challenge in this area is represented by an almost complete lack of official data on interregional trade. Accordingly, one line of research has focused on estimating bilateral trade between the regions of a country from existing data sources. These approaches are often cumbersome to implement, as they either rely on information that is noisy and incomplete (freight transport data) or require datasets that are only available under special circumstances (e.g. ad-hoc surveys, pre-existing regional IO tables) (TÃ¶bben 2017, Zheng at al. 2022).
Conversely, a large family of (â€˜non-surveyâ€™) techniques attempt to construct single-region IO models without having to estimate bilateral trade flows. For example, location quotient (LQ) methods derive the IO coefficients of a regional model by tweaking those of the country as whole. Even though their theoretical foundations and empirical performance have sometimes been called into question (Hermannsson 2016, Lamonica and Chelli 2018), LQs are very popular in applications thanks to their computational simplicity and minimal data requirements (BuendÃa et al. 2022, Flegg et al. 2021, Kwon and Choi 2023).

Another popular non-survey approach builds on the fundamental fact that in an input-output table total supply must match total use. Hence, commodity balance (CB) methods seek to work out interregional trade as a balancing item after all other components of the regional IO table (e.g., output, intermediate and final use) have been estimated (e.g. combining national IO and regional accounting data). In terms of minimum data requirements CB techniques are comparable with LQs, but they provide a more intuitive way of incorporating any additional information that may be available. The main drawback is that they yield estimates of net trade but cannot distinguish between imports and exports, although some workarounds have been proposed (Kronenberg 2009).

As many modern applications of IO analysis require multiregional models (e.g. greenhouse gas emission accounting and trade in value added analysis), non-survey regionalization methods are increasingly being generalized to accommodate multiple regions. Still, this tends to require additional assumptions whose relationship with the original single-region setup is not always transparent. Also, ensuring accounting consistency across regions often involves additional balancing steps, as a result of which these techniques are no longer as convenient as their original single-region version (TÃ¶bben and Kronenberg 2015, Jahn 2017).

This paper reports on the authorâ€™s efforts to develop a general procedure for constructing multiregional IO databases at the subnational level that would score well on the following three broad criteria. Firstly, it should rely on data that are routinely available from national statistical offices (e.g. national and regional accounts) but be flexible enough to accommodate any relevant additional data. Secondly, it should be based on a coherent set of standard theoretical assumptions. Thirdly, to the extent possible it should retain the ease of use of traditional non-survey approaches.

In practice, our approach can be described as combining the CB method with a doubly constrained gravity (DCG) model of interregional trade. Effectively, the DCG model is used to blow up the net trade estimates computed by CB regionalization into a full set of bilateral flows, thus separating imports from exports.

We start by defining an origin-destination matrix of interregional trade for each product (or industry) of the IO accounts. Its entries are unobserved but assumed to follow a gravity equation. Under very general conditions, this trade matrix can be recovered from just three pieces of information: 1. Bilateral interregional distances; 2. the (product-or industry-specific) distance elasticity of trade; 3. the row and column totals of the trade matrix itself. If estimates of these quantities are available, the trade matrix can indeed be estimated via a standard application of the RAS algorithm (Cai 2021). In applications, distance data are easy to obtain. Also, theory-backed distance elasticity estimates can be constructed using standard econometrics on widely available datasets, e.g., international trade data (Cai 2022). As for the last piece of the puzzle â€“ the trade matrixâ€™s marginal totals â€“ this paper shows that they emerge naturally when the CB method is applied region by region. Besides, those trade flow estimates balance the multi-regional system by construction, so that there is no need for further balancing.

The approach is demonstrated using a case study of Italy.