Title: Completing sub-national coverage of household accounts

Corresponding author: Roberto Zelli
Address: Faculty of Statistics, Sapienza University of Rome,
P.le Aldo Moro 5--00185 Rome, Italy
roberto.zelli@uniroma1.it,

Co-authors:
Nicholas T. Longford, SNTL and Universitat Pompeu Fabra, Barcelona, Spain and
Maria Grazia Pittau Sapienza University of Rome, Italy.

The focal problem of the proposed paper can be formulated as
follows.
The domain (a country or a region) is divided into several subdomains (provinces or districts). For a particular variable, say a national account variable, only its total over the entire domain is known, but the within-subdomain totals are of interest. Some information is available about the subdomains, such as their (population) size and some socio-economic indicators. This auxiliary information may come from administrative and/or survey data. Information about the within-subdomain totals may be also available from the past.

The methodology we propose in this paper seeks to draw on all this information by estimating the split of the domain-level total in a way that best conforms with the known split for the indicators. A distintive feature of our method is that the uncertainty about the estimates is represented by a small number of plausible solutions, with which a range of alternative scenarios can be explored.
The key technical device in the solution is evaluation of conditional expectations under the assumption of normality and a pattern of dependence of the subdomain totals.
This pattern may take into account the neighbourhood (spatial) structure of the subdomains (Anselin, 1988). The estimation will be accompanied by a sensitivity analysis in which we explore whether realistic departures from the adopted assumptions result in substantial changes in the estimates.

The outlined problem was first addressed by Chow and Lin (1971) in the context of time series: e.g., the total of a variable is known for the whole year, and we wish to estimate the totals within the quarters. Bollino (1998) and Polasek and Sellner (2008) pointed out that the quarters in the time series and subdomains of a domain present similar problems; they differ only in the association structure (time series vs. spatial dependence) of the units of analysis. The solution proposed by Polasek and Sellner and by Llano et al. (2009), in a Bayesian setting, is limited to auxiliary information with a restricted format, as is the original solution by Chow and Lin. Instead of the regression framework they apply, we will estimate the division from conditional expectations of the subtotals given the total and the auxiliary information, and will represent the uncertainty about the estimates by a set of plausible solutions.

Availability of coherent databases of national accounts indicators at different spatial disaggregation level (regional, provincial, local data) is particularly relevant in the European Union. In fact criteria for assignment and evaluation of European regional funds are essentially based on the availability

of timely and reliable regional accounts aggregates. These figures are collected in the European Union by Eurostat at different spatial levels according to the Nomenclature of Territorial Units for Statistics (NUTS). Even though in 2003 the Nomenclature was established as a common classification of territorial units for statistics, Regio (the Eurostat regional database) displays missing data for some countries and some variables at the lowest levels of territorial disaggregation.

In this paper we focus on the disaggregation of annual households accounts data, estimated according to the ESA95, at regional (NUTS2) and provincial (NUTS3) levels in Italy. The main source of micro data we want to use in our model is the EU Survey on Income and Living Conditions (EU-SILC). The Italian National Institute of Statistics (ISTAT) releases regional (NUTS2) households accounts with a delay of 24 months after the end of the reference period (e.g. regional households accounts for the period 2001--2006 were released by ISTAT on February 13th 2009), in line with the EU transmission programme of national accounts data of Regulation EC n. 1392/2007. Currently, ISTAT does not provide households aggregates at NUTS3 level.

We start estimating households accounts data at NUTS2 level in order to have more timely estimates, coherently with the national households accounts released by ISTAT, that represent in this step the domain-level total. The delayed estimates of regional households accounts coming from ISTAT will help us to evaluate our model. In the second step, we estimate aggregates of the households accounts at provincial (NUTS3) level, according to the proposed model. In this analysis we will focus on the estimation of total and disposable household income and its components. Our methodology can be easily adapted to other European countries and other aggregates.

References

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Kluwer Academic Publisher, Dordrecht.

Bollino, A. (1998). Econometric Interpolation of Regional Time Series, Statistica Applicata (Journal of Applied Statistics), n.10

Chow, G.-C., and Lin, A. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. The Review of Economic and Statistics 53, 372–375.

Llano, C., Polasek, W., and Sellner, R. (2009). Bayesian methods for completing data in space-time panel models. Institute for advanced studies, Vienna. Economics Series n. 241.

Polasek, W., and Sellner, R. (2008). Spatial Chow-Lin methods: Bayesian and ML forecast comparisons. The Rimini Centre for Economic Analysis, WP38–08.