BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Timely quarterly indicators of household consumption and disposable income for the Italian regions

by Valter Di Giacinto, Vincenzo Mariani, Daniele Ruggeri, Giuseppe Saporito, Andrea Sechi, Giovanni Soggia, Andrea Venturini and Antonio Veronico

October 2025

Number 970

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Timely quarterly indicators of household consumption and disposable income for the Italian regions

by Valter Di Giacinto, Vincenzo Mariani, Daniele Ruggeri, Giuseppe Saporito, Andrea Sechi, Giovanni Soggia, Andrea Venturini and Antonio Veronico

Number 970 – October 2025

The series Occasional Papers *presents studies and documents on issues pertaining to the institutional tasks of the Bank of Italy and the Eurosystem. The* Occasional Papers *appear alongside the* Working Papers *series which are specifically aimed at providing original contributions to economic research.*

The Occasional Papers *include studies conducted within the Bank of Italy, sometimes in cooperation with the Eurosystem or other institutions. The views expressed in the studies are those of the authors and do not involve the responsibility of the institutions to which they belong.*

The series is available online at www.bancaditalia.it .

ISSN 1972-6643 (online)

Designed by the Printing and Publishing Division of the Bank of Italy

TIMELY QUARTERLY INDICATORS OF HOUSEHOLD CONSUMPTION AND DISPOSABLE INCOME FOR THE ITALIAN REGIONS

by Valter Di Giacinto*, Vincenzo Mariani**, Daniele Ruggeri◊, Giuseppe Saporito◊◊, Andrea Sechi†, Giovanni Soggia†, Andrea Venturini†† and Antonio Veronico‡

Abstract

This study introduces an innovative econometric methodology for developing timely quarterly indicators of household income and consumption for all Italian regions and autonomous provinces. Leveraging a comprehensive statistical database spanning from 1995 to 2022, the methodology utilizes basic indicators from the real economy and monetary sectors. These indicators are condensed into regional common factors and combined with national aggregates to model the annual regional time series published by Istat. Two selection approaches, Stepwise Forward Selection (SFS) and sparse Temporal Disaggregation (spTD) using LASSO-type methods identify relevant local factors. These models interpolate observed annual figures *ex-post* and provide *ex-ante* estimates of quarterly regional aggregates with a 90-day post-quarter-end lag. In-sample evaluations show high model fit, particularly with the spTD methodology. Out-of-sample forecasting confirms satisfactory predictive performance, albeit with varying precision across regions. Overall, this methodology yields reliable indicators of regional household income and consumption dynamics, offering timely insights that are crucial for short-term economic analysis and answer the challenges posed by official statistics in Italy, which are annual and released with a one-year lag.

JEL Classification: C22, C82, E01.

Keywords: temporal disaggregation, regional economies benchmarking and extrapolation, real time series.

DOI: 10.32057/0.QEF.2025.970

* Bank of Italy, L'Aquila Branch; corresponding author: valter.digiacinto@bancaditalia.it.

** Bank of Italy, Bari Branch; corresponding author: vincenzo.mariani@bancaditalia.it.

◊ Bank of Italy, Rome Branch.

◊◊ Bank of Italy, Palermo Branch.

† Bank of Italy, Cagliari Branch.

†† Bank of Italy, Venice Branch.

‡ Bank of Italy, Bari Branch.

1 Introduction[1]

Household consumption and income are key indicators of the well-being of a population, as they approximate individual utility and the ability to acquire goods and services. They provide valuable information that can assist policymakers in designing more suitable and targeted policies.

In Italy, official statistics on regional consumption and income are produced annually by the National Institute of Statistics (Istat) as part of the regional accounts (*Conti economici territoriali*). However, these regional accounts are published with significant delay: preliminary estimates of consumption and disposable income for resident households are typically available about one year after the reference period. These estimates, along with other aggregates, undergo revisions over subsequent years as more data becomes available. The delayed availability of these statistics limits their utility for providing a real-time assessment of the regional juncture, particularly in a country like Italy characterized by significant disparities across regions and different economic cycles.

Furthermore, unlike national data, which is available on a quarterly basis, the official regional data is aggregated annually. This disparity makes it impractical to conduct cyclical analyses of household consumption and income patterns at smaller geographic levels.

This research aims to address these limitations by introducing an innovative methodological approach to timely estimate income and consumption at the regional level on a quarterly basis. Using advanced econometric models and detailed analysis of available data, we contribute to the economic debate by offering a more detailed perspective on economic dynamics and the evolution of well-being across Italian regions.

The methodology employed draws inspiration from the real-time GDP estimation approach used by Di Giacinto et al. (2019) for Italian macro-regions, as well as aspects of the work by Cuevas, Quilis, and Espasa (2015), which proposed a similar methodology for producing timely quarterly GDP estimates for Spanish regions.

In terms of data requirements, the procedure relies on three main components: (a) annual series of household disposable income and consumption expenditure at the regional level, (b) equivalent quarterly national time series, and (c) approximately 30 quarterly regional indicators correlated with local trends in income and consumption. These regional indicators encompass both real variables (e.g., local labor market measures) and financial variables (e.g., credit card payments and household loans) spanning nearly 30 years from 1995 to 2023.

[1] The views expressed in this paper are those of the authors and do not involve the responsibility of the Bank of Italy and/or the Eurosystem. We thank Giuseppe Albanese, Valentina Aprigliano, Marta Crispino, Silvia Del Prete, Marcello Pagnini, Concetta Rondinelli, Roberto Torrini and Giordano Zevi for fruitful discussions and the participants to seminars at the Bank of Italy for comments and suggestions. We also thank Irene Longhi and Nazzareno Renzi for the provision of UIF-SARA aggregated regional statistics and Luca Bastianelli for the provision of credit card payment data. All errors are our responsibility.

The econometric procedure allows for the estimation of updated quarterly regional series within approximately 90 days of each quarter's end. These series are designed to be consistent both in their annual sum with corresponding official data and in their sum across regions with national quarterly data.

A key advantage of this procedure, compared to the economic cycle indicator ITER proposed by Di Giacinto et al. (2019), is its ability to provide coherent, efficient, and centralized estimates for all Italian regions by leveraging the strong correlation between national and regional income and consumption dynamics.

In Italy, some organizations produce regional estimates of consumption and disposable income, albeit only annually. For instance, Svimez (an independent, non-profit research institute), and Confcommercio (an association representing businesses in the service sector), publishes local annual consumption indicators at year-end. Similarly, Prometeia, an Italian consulting and economic analysis company, develops annual indicators for household consumption and income, available for purchase by businesses and local administrations.

The structure of this work is as follows: Section 2 outlines the econometric procedure, detailing how regional and annual input time series are used to generate quarterly regional indicators. Section 3 provides a more detailed description of the input series, including explanations for any reconstructed series in years where data was not initially available. Section 4 presents evidence on consumption and income outcome series. Section 5 discusses estimation results, illustrates the goodness of fit of various models, and the prediction error compared to official statistics. Finally, Section 6 summarizes the main conclusions of the study.

2 Methodology

2.1 The quarterly temporal disaggregation model

We start by assuming that the unobserved quarterly regional series y_{rt} of the level of households' income (consumption) at constant prices is represented by the following linear common factor model

$$y_{rt} = \mu_r + \alpha_r G_t + \eta_{rt} \qquad (1)$$

where $r = 1,2,\dots,R$ and $t = 1,2,\dots,T$ denote the region and the quarter respectively; the variable μ_r is a constant that captures time-invariant level differences between regions, G_t is a factor common to all regions, and η_{rt} is a regional specific component, possibly correlated with G_t and weakly cross-section correlated. Both the common national and the regional components are assumed to be integrated of order 1 or 0.

We also assume that the regional idiosyncratic component has a factor structure of the following type

$$\eta_{rt} = \beta_r' F_{rt} + u_{rt} \qquad (2)$$

where F_{rt} is a vector of K factors specific to the r-th region and u_{rt} is a stochastic error term, orthogonal to F_{rt}, which is assumed to evolve according to the following first order autoregressive process $u_{rt} = \rho_r\, u_{rt-1} + \varepsilon_{rt}$. The error term ε_{rt} is white-noise and has constant variance σ_r^2. Also in this case, both the factors that make up F_{rt} and the stochastic disturbance term u_{rt} can be I(1) or I(0).

Substituting (2) into (1) we obtain the following equation, which constitutes the model on which the construction of the quarterly income and consumption indicators is based on

$$y_{rt} = \mu_r + \alpha_r G_t + \beta_r' F_{rt} + u_{rt} \qquad (3)$$

Equation (3) makes explicit how, under our assumptions, each outcome variable (in a given quarter and for a given region) is the sum of a constant, a common national factor, a few regional factors and an autoregressive error term.

Given that the outcome is observed only on an annual basis, in order to obtain an equation that can be estimated from the data, we rewrite (3) by aggregating temporally, adding the four quarterly observations of each year. We obtain the following expressions

$$\boldsymbol{y}_r = \mu_r + \alpha_r \boldsymbol{G} + \boldsymbol{F}_r \beta_r + \boldsymbol{u}_r \qquad (4)$$

$$\overline{\boldsymbol{y}}_r = \overline{\mu}_r + \alpha_r \overline{\boldsymbol{G}} + \overline{\boldsymbol{F}}_r \beta_r + \overline{\boldsymbol{u}}_r \qquad (5)$$

where $\boldsymbol{y}_r = [y_{r1}, y_{r2}, \dots, y_{rT}]'$, $\boldsymbol{G} = [G_1, G_2, \dots, G_T]'$, $\boldsymbol{u}_r = [u_{r1}, u_{r2}, \dots, u_{rT}]'$,

$$\boldsymbol{F}_r = \begin{bmatrix} \mathrm{F}_{r1}' \\ \mathrm{F}_{r2}' \\ \dots \\ \mathrm{F}_{rT}' \end{bmatrix}$$

and where $\overline{\boldsymbol{y}}_r = C\boldsymbol{y}_i$ coincides with the observed regional series of household income (consumption) and where $\overline{\mu}_r = C\mu_r$, $\overline{\boldsymbol{G}} = C\boldsymbol{G}$, $\overline{\boldsymbol{F}}_r = C\boldsymbol{F}_r$, $\overline{\boldsymbol{u}}_r = C\boldsymbol{u}_r$, with C denoting the temporal aggregation matrix $C = (I \otimes c)$, where $c = [1\ 1\ 1\ 1]$.

At this point, to make (5) estimable, it is necessary to assign values for the series of national and regional common factors. To this end, in the implementation proposed here, we use the corresponding national aggregate of household income (consumption), denoted as y_{Nat} hereafter, as a proxy for the common factor G. Under the hypothesis that the idiosyncratic regional components of income and consumption are only weakly cross-sectionally correlated, aggregation at the national level tends to cancel them, allowing only the common component to emerge.

On the contrary, to identify and estimate the regional common factors, we resort to the procedure proposed by Bai and Ng (2004, BN hereafter). This procedure does not require the stationarity of the observed series and allows us to extract common factors that can present unit roots, a characteristic notoriously of many macroeconomic series, income and consumption included. Following the approach by BN, we therefore assume that the local common factors $\boldsymbol{F}_r$ are linked to a set of regional indicators X_{rit}, $i = 1, \dots, M_r > K$, by the following static linear relationship

$$X_{rit} = c_r + \lambda'_{ri}\boldsymbol{F}_{rt} + e_{rit} \quad (6)$$

where both the factors $\boldsymbol{F}_{rt}$ and the idiosyncratic components e_{rit} are allowed to be I(1) or I(0).

Under the assumptions given in BN, it can be shown that a consistent estimator of the K common factors can be obtained by applying the principal components method to the data taken in first differences. In particular, the expression of the estimator proposed by BN is the following

$$\hat{F}_{rt} = \sum_{s=2}^{t} \hat{f}_{rt} \quad (7)$$

where $\hat{f}_{rt}$ is proportional to the eigenvectors of the matrix $x_r x_r'$, of dimensions $(T-1) \times (T-1)$, where $x_r = [x_{r1}, x_{r2}, \dots x_{rM_r}]$, with $x_{ri} = [\Delta X_{ri2}, \Delta X_{ri3}, \dots \Delta X_{riT}]'$, $i = 1, \dots, M_r$.

We subsequently stack observations for y_{Nat} and the estimated factors $\hat{F}_{rt}$ and replace the unknown coefficients in equation (4) with the corresponding maximum likelihood estimators. Imposing the temporal aggregation constraint

$$\sum_{q=1}^{4} \hat{y}_{raq} = \bar{y}_{ra} \quad (8)$$

where $\hat{y}_{raq}$ denotes the quarterly disaggregated series for region r, year a and quarter q, and where $\bar{y}_{ra}$ is the corresponding observed annual series, as shown in Chow-Lin (1971) the optimal estimator of the targeted quarterly series is

$$\hat{\boldsymbol{y}}_r = \bar{\hat{\mu}}_r + \hat{\alpha}_r \boldsymbol{y}_{Nat} + \hat{\boldsymbol{F}}_r \hat{\beta}_r + \hat{\Omega}_r C' \left(C \hat{\Omega}_r C'\right)^{-1} \left(\bar{\boldsymbol{y}}_r - \bar{\hat{\mu}}_r + \hat{\alpha}_r \bar{\boldsymbol{y}}_{Nat} + \bar{\hat{\boldsymbol{F}}}_r \hat{\beta}_r\right) \quad (9)$$

where $\hat{\Omega}_r$ is the maximum likelihood estimator of error covariance matrix $\Omega_r = E(\boldsymbol{u}_r \boldsymbol{u}_r')$, and where $\bar{\boldsymbol{y}}_{Nat} = C\boldsymbol{y}_{Nat}$ and $\bar{\hat{\boldsymbol{F}}}_r = C\hat{\boldsymbol{F}}_r$.

This formula allows the interpolation/distribution ex-post (i.e. when the annual total is known) of the corresponding annual historical series. However, the temporal disaggregation approach can also be extended to out-of-sample forecasts, or ex-ante estimates, of the quarterly series, to be produced before the publication of the corresponding annual series. By exploiting the information on high-frequency indicators that becomes available during the year, it is indeed possible to extrapolate the data for the quarterly series. The ex-ante estimates, for $t = T+1, T+2, \dots, T+H$ can be computed using the following recursive procedure:

$$\hat{\hat{y}}_{i,T+h} = \hat{\mu}_r + \hat{\alpha}_r G_{T+h} + \hat{\beta}_r' \hat{F}_{rT+h} + \hat{u}_{i,T+h} \quad (12)$$

where

$$\hat{u}_{i,T+h} = \hat{\rho}\hat{u}_{i,T+h-1}, \quad h = 1,2, \dots, H. \quad (13)$$

2.2 The selection of the regional common factors

Not all the common factors estimated by the BN procedure will actually contain information useful in order to predict the local dynamics of household income and consumption. Consequently, in order to achieve parsimony and efficiency in model building, we implement a procedure to select the subset of the most significant regional factors.

For this purpose, we consider two alternative approaches: a Stepwise Forward Selection (SFS) procedure, and the sparse temporal disaggregation (spTD) method, recently proposed by Mosley, Eckley and Gibberd (2022a, MEG in what follows), that applies the LASSO regression to the traditional Chow-Lin modeling environment.[2]

In the SFS procedure we consider up to a maximum of three local factors for each region; we select the factors with the greatest explanatory capacity in the quarterly disaggregation model, in addition to the national income/consumption series. In the appendix we detail the procedure implemented.

2.3 The spatial aggregation constraint and the procedure to balance the quarterly series

In the system of annual national economic accounts, the regional series satisfy the constraint that the sum of income and consumption values of all the regions is identically equal to the corresponding national aggregate. It is therefore appropriate to require that quarterly estimates of regional income and consumption series also satisfy the following spatial aggregation constraint

$$\sum_{r=1}^{R} y_{rt} = Y_t \quad (14)$$

where Y_t denotes the corresponding national aggregate.

Following the approach of Cuevas et al. (2015) and Di Giacinto et al. (2019) the transversal aggregation constraint (14) is imposed in a second stage, after having obtained the quarterly series for each region in the first stage using the procedure described in the previous paragraph. In particular, in the ex-post balancing of the quarterly national series, we implement the biproportional algorithm of Bacarach (1965), which guarantees, through an iterative optimization technique, that in each year both the temporal and transversal constraints are respected.

[2] Estimations have been performed using the R package described in Mosley et al. (2022b).

Before applying the two-dimensional balancing algorithm, the raw series produced by the first stage estimates are subjected to a smoothing and shrinkage procedure which reduces their dispersion in the temporal and spatial dimension, having as a benchmark the volatility of the national income and consumption and the cross-section dispersion of the rates of change of the corresponding regional series.

The temporal smoothing of the series is obtained by taking weighted centered moving averages of three terms of the series $\hat{y}_{rt}$, with weights equal to [1/6, 2/3, 1/6]. A shrinkage of the trend rates of change of the quarterly series with respect to the dynamics of the corresponding national quarterly series is subsequently applied to the smoothed series, which are imposed to satisfy the following relationship

$$(\hat{y}^{*}_{rt} - \hat{y}^{*}_{rt-4})/\hat{y}^{*}_{rt-4} = (1 - S)(\hat{y}_{rt} - \hat{y}_{rt-4})/\hat{y}_{rt-4} + S(Y_{rt} - Y_{rt-4})/Y_{rt-4} \qquad (15)$$

where $\hat{y}^{*}_{rt}$ is the value of the quarterly series after the shrinkage and S is a coefficient between 0 and 1 which measures the amount of shrinkage applied to the dynamics of the regional series compared to the dynamics of the corresponding national series. This operation brings the dynamics of the individual regional series closer to that of the national series. In addition, containing the cross-section dispersion of the estimates, produces a pre-balancing of the latter, reducing the discrepancy with respect to the national aggregate. The remaining discrepancy is subsequently distributed between the regions so as to ensure that the transversal aggregation constraint (10) is verified in each quarter.

In the case of ex-ante estimates, since the corresponding annual data is not yet available, the balancing of the quarterly regional series with the corresponding national series, once the smoothing and shrinkage procedures have been applied, is obtained through a simple procedure of proportional distribution of the discrepancy, similar to that considered in Proietti (2002) and Di Giacinto et al. (2019). Through this procedure, in each quarter the difference between the national series and the sum of the values of the regional series is attributed to the latter in proportion to the size of the reference aggregate (income or consumption).

3 The data

This section describes the indicators utilized in order to portray the national and the local idiosyncratic dynamics of household income and consumption at the regional level. The series (outcomes or variables that, according to the theory, can be useful for their prediction) cover the period from 1995 to 2022: as in some cases the series were not complete, we describe how we imputed missing observations.

1. Real indicators:

a) Household income and consumption at national and regional level (Source: *Conti nazionali* and *Conti economici territoriali*, Istat) – Istat produces estimates of the national and regional accounts which are compliant with ESA 2010 (the European system of national accounts).[3] The national account system describes the economic activity of a territory (either at national or a sub-national level) and the economic and financial flows between the economic operators during a period, either one year or a sub-period. Istat estimates national and regional aggregates by combining several data sources ranging from surveys to administrative information (Istat, https://www.istat.it/en/archivio/115500). From the edition released on January 2024, we collected quarterly seasonally-adjusted national data of gross disposable income of consumer households at current prices, from 1998 to 2023[4]. From the edition released on December 2023 we collected quarterly seasonally-adjusted data of the final consumption expenditure of resident and non-resident households on the economic territory, both at current prices and at chained linked values with 2015 as reference year, over the period 1995-2023. We imputed the four missing years of gross disposable income (from 1995-Q1 to 1998-Q4) by assuming that it followed the same seasonally-adjusted dynamics of compensation of employees, which is also available from the national accounts. Then, we expressed Italy's gross disposable income in real terms by using the implicit deflator that we drew from the national final consumption expenditure series. From December 2023 edition of regional accounts we collected the same aforementioned series at yearly frequency (from 1995 to 2022) for all Italian regions; we expressed gross disposable income of regions in real terms using the national deflator described above.

b) Labor force survey data (Source: *Rilevazione continua sulle forze di lavoro*, Istat) – We use quarterly data from the Labor force survey at the regional and national level, which encompasses various labor market indicators, all referred to the resident population. These indicators include: the number of employed and unemployed, the labor force, the number of individuals outside the labor force, the total population, the number of self-employed and employees, the activity, employment and unemployment rates, the average weekly hours worked by an employed person and the total number of hours worked. Most of the data pertain to the population aged 15 and older, except for the activity and employment rates, which refer to those aged 15 to 64, and the unemployment rate (15-74). All these series exhibit a structural break in 2018, when the rules for classifying employment changed, in line with Regulation (EU) 2019/1700. For some regions (Bolzano, Trento, and Valle d'Aosta), data on the average weekly hours worked were available only since 2009.

[3] Regional accounts refer to the 19 Italian regions and to the two autonomous provinces (Trento and Bolzano). We consider the same classification in the rest of this paper.

[4] We exclude, in particular, sole proprietorships and partnerships without legal status (*famiglie produttrici*) and non-profit institutions serving households (*istituzioni sociali private*).

For the missing years (1995-2008), we imputed hours worked by assuming that the series followed the same dynamics as in the corresponding macro-area[5].

c) Consumer confidence survey data (Source: *Indagine sulla fiducia dei consumatori e delle imprese*, Istat) – We use seasonally-adjusted data from the consumer survey, conducted monthly by Istat since 1982. The survey reports the evaluation on the economic situation of the respondent and on the broader economic conditions. The consumer confidence index is calculated as the weighted average of responses to nine questions, expressed as balances. Each balance is expressed as the percentage difference between the share of respondents who reports an improvement and that of respondents who reports a worsening. Together with the main confidence index, we also use other four indexes, which are calculated by Istat using subsets of the initial nine questions (current/future economic sentiment; personal/general economic sentiment). These indexes are available only for Italy and the macro-areas, hence we assign to each region the value of its macro-area. The first available year is 1998; given that we consider the period from 1995 to 2022, we filled the data for the three missing years by assuming for the series of the macro-areas the same dynamics of the national indicator, which is available between 1995 and 1997. The quarterly data is obtained by averaging the corresponding monthly series.

d) Residential transactions. (Source: *Agenzia delle Entrate, Osservatorio sul mercato immobiliare*) – We consider the number of residential transactions registered in each quarter in the regional real estate market, as provided by the Italian Tax Agency.

e) Tourist overnight stays (Source: *Movimento dei clienti negli esercizi ricettivi*, Istat) – This series reports the number of nights spent in accommodation establishments (hotels and other types) every month and in each region[6]. The inclusion of this series is motivated by the fact that one of the outcome, consumption, refers to the expenditure on the territory, irrespective of the residence.

f) New car registrations (Source: ANFIA) – This series reports the monthly number of new cars registered at the regional level. Car registrations are a component of consumption of durable goods.

g) New business registrations (Source: Infocamere[7]) – This series, which is considered as a proxy of the local economic cycle, reports the number of new individual firms (sole proprietorships) recorded in each quarter in the general business register held by the Chambers of commerce.

2. Financial and payments indicators:

[5] The regions and the autonomous provinces are classified in macro-areas (NUTS1) as follows: Piedmont, Valle d'Aosta, Lombardy, Liguria (North-West); Veneto, Trentino, Bolzano, Friuli-Venezia Giulia, Emilia-Romagna (North-East); Tuscany, Umbria, Lazio, Marche (Center); Abruzzo, Molise, Campania, Apulia, Basilicata, Calabria, Sicily, Sardinia (South).

[6] See Crispino and Mariani (2023) for a description of these data.

[7] Infocamere is the statistical office of the system of the Italian Chambers of Commerce.

a) Aggregate anti-money-laundering statistics (SARA) (Source: Financial Intelligence Unit, UIF). – The Italian anti-money-laundering legislation requires that banks and other financial intermediaries report monthly to UIF all the financial transactions exceeding a given threshold. AML statistics, once properly treated in order to reduce their inherent volatility and to control for the modifications of the reporting threshold that have occurred over time, have proved to possess appreciable predictive power for the growth rate of regional GDP (Aprigliano et al., 2021). Aggregate regional time series for the total flows, cash flows, domestic bank transfers and foreign bank transfers, separately for debit and credit figures, were considered. The individual series were pre-processed by the TRAMO-SEATS procedure in order to remove outliers and level breaks and seasonally adjust the data. Given that the SARA series are available only from year 2001 onwards, the missing values for the period 1995Q1-2000Q4 were imputed by means of the fitted values of the regression of each regional series on the five regional factors most correlated with the former. Such factors are estimated by the Bai and Ng (2004) method taking as inputs all the remaining regional indicators in the dataset described in this Section.

b) Value of point of sales (POS) transactions (Source: Nexi Payments) – In order to capture consumers' expenditures made through card payments, we follow existing studies (Crispino and Mariani, 2023; Emiliozzi, Rondinelli and Villa, 2023) and use data on the international circuit debit cards (Mastercard and Visa) managed by Nexi Payments: these data refer to the transactions made by cards at the points of sales of merchants located in Italy (so-called acquiring side), including both physical and virtual points of sales. The anonymized daily information has been collected since May 2014 and contains details on the region where the payment was made; the dataset also distinguishes 11 product categories, ranging from clothing and food to hotels and restaurants. For the period from 1999Q4 to 2014Q1, the series were reconstructed by applying the dynamics of the number of credit cards held in each region by non-business owners (Source: Bank of Italy, Supervisory reports). The half-yearly baseline data were prior linearly interpolated at a quarterly frequency. For the period 1995-1999, statistics on the amount of credit card payments were used for the reconstruction of the regional payments series. The data utilized for this period were published in the annual report of the Bank of Italy for the year 2000, with territorial detail for the Northern, Central and Southern Italian geographical areas. Also in this case, quarterly figures were obtained by interpolating linearly the annual figures.

c) Interest rates (Source: ECB) – We consider the Euribor 3-month interest rate, averaging the monthly observations through each quarter.

d) Households financial indicators (Source: Bank of Italy). – The data are obtained from the periodical supervisory reports that banks and other financial intermediaries submit to the Bank of Italy as mandated by the law. In particular, we use two different series: a) the outstanding amounts of total bank loans to households; b) the outstanding amounts of consumer credit by banks and financial institutions to households. In both cases, the data are allocated to each region according to the household's place of residence. Given that the banks consumer credit series are available from 2002 onwards, for the missing years (1995-2001) we imputed the missing data by assuming that the series followed the same dynamics as the series of consumer credit granted by financial institutions.

e) Consumer price index (Source: Istat). – We consider the consumer price index for the entire national community (NIC). Data are available for every region with monthly frequency. For 1995, as the index is not available, we rely on the Consumer Price Index for blue-collar and white-collar families (F.O.I.) net of tobacco.

4 Preliminary evidence on the outcome variables

As already mentioned, to track the regional dynamics of income and consumption, we employ: gross disposable income of resident consumer households and final consumption expenditure of resident and non-resident households on the economic territory. Time series for the two aggregates were collected both at the national and regional level and cover the period 1995-2023 at the quarter (for national aggregates) and year frequencies (for regional aggregates). For national quarterly series, we retrieved seasonally-adjusted data as distributed by Istat. Both for Italy and the regions, gross disposable income data are only available at current prices, while the final consumption series is available both at current prices and at chain-linked values, using 2015 as reference year. Consequently, both for Italy and all the regions, we expressed, as already mentioned, gross disposable income in real terms by applying the implicit deflator of the national final consumption series. To carry out our analyses we rely on the latest release of the data, i.e. December 2023 for the national's quarterly final consumption and for regions' yearly final consumption and gross disposable income; data from January 2024 edition were employed for the quarterly series of national gross disposable income.

Figures 1 and 2 respectively plot the time series of the annual aggregates of income and consumption for each region and Italy. The visual inspection immediately reveals a strong comovement between the regional and the national aggregates, with a general coincidence of the business cycle turning points. The series also clearly appear to be non stationary, a preliminary graphical evidence that is confirmed by the results of individual unit root tests, that, apart from some marginal exceptions, prove the series to be I(1), i.e. non stationary in levels but stationary in first-differences (see Table 1).

Table 2 displays the bivariate correlations between the regional income and consumption series and the corresponding national totals. The correlations are generally quite high, both when the variables are measured on levels and on first-differenced data, providing thus evidence that the correlations are not spurious but may reflect the presence of an underlying common stochastic trend, contributing to drive the temporal evolution of the phenomena across all the areas. Under the assumption that only the common national trend drives the long run dynamics of the regional series, the regional and the national series should be cointegrated. To verify this hypothesis, we run a series of bivariate Engle-Granger cointegration tests, whose results are also given in Table 2. The null hypothesis of the absence of cointegration cannot be rejected for almost all the 44 cases considered. The evidence of cointegration tests is thus strongly in favour of the existence of local idiosyncratic trends that contribute to determine the long run evolution of the regional aggregates in combination with a common national trend.

5 Estimation results

The first step in the estimation of the quarterly regional income and consumption indicators involves the estimation of regional common factors. To this purpose, as preliminary treatment, the time series of the indicators of local economic conditions were differenced, seasonally adjusted (by regressing them on a set of seasonal dummies) and standardized. The BN estimators are subsequently applied to the pre-treated data, yielding a set of estimated common factors. Figures 3 and 4 display the first three common factors for each region, selected according to the forward stepwise procedure outlined in Section 2.2. In general, different factors are selected in any given region for the income and consumption series (a maximum of one shared factor between the two series occurring in approximately half of the cases).

The graphical analysis clearly shows that some of the factors are trending and non-stationary. The latter feature can therefore be deemed to possess some potential in capturing the local idiosyncratic stochastic trends that, as the preliminary evidence shows, drives the local dynamics of household income and consumption together with a common national trend.

Utilizing the series of the regional factors yielded by the procedure prior outlined, the model given by equation (8) was subsequently fitted separately for each region and each dependent variable (income and consumption). Tables 3 and 4 display, separately for the income and consumption series, some baseline measures of in-sample goodness-of-fit for the individual model specifications, together with value of the ρ-parameter, gauging the degree of serial correlation of the model residuals (a value of ρ approximately equal to 1 provides evidence of the absence of cointegration between the dependent variable and the set of national and local factors that enter as predictors in the temporal disaggregation model equation).

In order to provide some evidence on the relevance of the common national factor, as a first specification, we estimate a model including only the corresponding national quarterly series. The estimation results show how the common national trend possesses high explanatory power in all the regional models here considered. Model specifications adding from 1 to 3 local regional factors, selected on the basis of the SFS methodology outlined above, were subsequently estimated. The fit of the regional models improves significantly with the first regional factor included, and up to the second, while only a marginal increase in the correlation between the dependent variable and the values predicted by the regression function is observed when the third factor is included. This provides evidence that the further addition of a fourth local factors is generally not required in order to improve the model fit. Finally, the in-sample predictive performance of the models is satisfactory not only with respect to the series in levels but also when y-o-y rates of change are considered.

Apart from a limited number of cases, the estimated ρ parameter is substantially smaller than 1. While not providing a formal cointegration test, this evidence appears to support the hypothesis that the national and regional factors utilized as predictors in temporal disaggregation models possess some capacity to capture the regional idiosyncratic trends, whose existence was prior uncovered by the cointegration analysis between the annual regional income/consumption series and the corresponding national aggregate carried out in Section 4.

In Tables 3 and 4 the goodness-of-fit statistics for the spTD and the adaptive-spTD model specifications are also displayed. For almost all the regional models the predictive ability shows an appreciable improvement when moving to spTD specifications, compared to the specifications yielded by the SFS factor selection procedure. Figure 5 (6) reports for each region the value of the indicator for disposable income (consumption) obtained from the model with the best predictive ability.

The Y-Xb correlation is now very close to unity in most cases and the estimated ρ shows a further decrease, being very close to zero in some cases. This results are not surprising, considering that the LASSO methodology underlying the spTD approach exploits all the regional factors extracted by the BN procedure for model fitting purposes, albeit with penalized regression coefficients, while the SFS approach here implemented favours models' parsimony in more stringent way.

6 The forecasting exercise

In order to gauge the out-of-sample predictive ability of the alternative model specifications, we carried out a forecasting exercise. For each year T ranging from 2016 to 2022, the forecasts of the annual aggregate of regional income and consumption were computed, on the basis of model parameter estimates obtained utilizing sample data up to the year T-1, by summing the values predicted by the temporal disaggregation model for the four quarters of year T.

The forecast of the annual growth rate of regional income/consumption was then compared to the ex-post growth rate computed on data for year T subsequently released by Istat and the absolute deviation between the predicted and the observed growth was computed. Tables 5 and 6 display the mean absolute forecast error (MAE) for the individual combinations of dependent variable, region and model specification, separately for the unbalanced forecasts and the forecasts balanced to the corresponding national aggregate.

A first baseline result of the forecasting exercise provides evidence that the use of the national anchor (balancing) in forecasting generally improves the model's predictive ability, as it lowers the forecast errors for both the regional income and consumption series. The models specified by means of the forward selection procedure stand up as the most accurate in the majority of the cases, with an average MAE across regions of 0.7 and 0.6 respectively for income and consumption. In particular, the specification with two local factors attains the lowest MAE in forecasting the household income series for all the largest regions located in the Center and the Northern areas of the country; for the consumption expenditure, the model with three local factors is found out to provide the most accurate forecasts in 10 regions out of 21, but without a clear territorial pattern in this case. Regional models specified according to the two variants of the sparse temporal disaggregation method provide better out-of-sample predictive ability in about a quarter of the regions considered in the exercise (5 for income and 6 for consumption).

When the best-performing model is selected for each region, the average MAE across regions is 0.7 for income and decreases to 0.5 for consumption. Figures 7 and 8 provide a graphical representation of the results, comparing the official statistics with the estimates year by year.

7 Summary and conclusion

In this work we implement an econometric methodology with some novel features, in order to develop timely quarterly indicators of household income and consumption for the entire set of Italian regions and autonomous provinces.

The new quarterly indicators exploit a rather rich and complex statistical database of basic indicators observed at quarterly frequency or higher, for which it was possible to reconstruct statistically continuous historical series for the entire period from 1995-Q1 to 2022-Q4. The basic indicators considered concern both aspects of the real economy and monetary and financial variables, which the literature suggests to be correlated with the dynamics of household income and expenditure.

The information conveyed by the basic regional indicators was initially condensed into a set of common factors with techniques that allow the treatment of time series in levels, generally well-known to be non-stationary due to the presence of unit roots.

The regional common factors, together with the corresponding national aggregate, included in order to capture trends common to all Italian regions, were subsequently utilized as explanatory variables in a battery of temporal disaggregation models of the annual regional time series of household income and consumption published by Istat.

To this end, two different approaches to the selection of the most relevant local factors were considered: the Stepwise Forward Selection (SFS) method and the sparse Temporal Disaggregation (spTD) method, which implements LASSO-type techniques in the context of traditional temporal disaggregation methodologies, used for many years by different national statistical institutes in the production of quarterly economic accounts.

By relying on the observed dynamics of the baseline indicators, the temporal disaggregation models, while allowing for the interpolation of the observed annual figures of regional household income and consumption expenditure (i.e. ex-post estimates), may also yield ex-ante estimates of the quarterly evolutions of the regional aggregates prior to the release of the annual estimates, with a delay of approximately 90 days from the end of the quarter.

The in-sample goodness-of-fit of the estimated regional models is generally highly satisfactory for all regions and for both type of modelling approaches, although some noticeable improvements are found out when the spTD methodology is implemented.

Considering that in the actual utilization of the indicators for short-term economic analysis ex-ante estimates that anticipate the release of the official annual figures will be generally utilized, the ability of the regional econometric models in correctly anticipating the annual dynamics of the household regional aggregates was subsequently evaluated by performing an out-of-sample forecasting exercise covering the 2016-2022 period.

The results of the forecasting exercise show an overall satisfactory predictive performance for both the income and consumption series. The forecast error is generally small for the larger regions of the Center and North areas, while, in line with expectations, the greater volatility of the annual series leads to a lower precision of the forecasts in the case of the smaller regions.

The evidence above documents how the methodological approach implemented in this study can be deemed to have yielded reliable indicators of the regional patterns of household income and consumption in Italy. By providing the analysts of local short-term economic evolutions with information available on a quarterly basis and updated in a timely fashion, these new indicators may help overcome the difficulties posed at the moment by the availability of official statistics only on annual basis and released with a delay of about one year from the end of the reference period.

References

Aprigliano, V., Ardizzi, G., Cassetta, A., Cavallero, A., Emiliozzi, S., Gambini, A., Renzi, N. and Zizza, R. (2021), "Exploiting payments to track Italian economic activity: the experience at Banca d'Italia", Banca d'Italia, Questioni di economia e finanza (Occasional Papers), N. 609.

Bacharach, M. (1965), "Estimating non-negative matrices from marginal data", International Economic Review, Vol. 6, n. 3, pp. 294-310.

Bai, J., Ng, S. (2004), "A PANIC attack on unit roots and cointegration", Econometrica, Vol. 72, n. 4, pp. 1127-1177.

Belloni, A. and Chernozhukov, V. (2013). "Least squares after model selection in high-dimensional sparse models." Bernoulli 19(2), 521-547.

Cuevas, A., Quilis, M., Espasa, A. (2015), "Quarterly regional GDP flash estimates by means of benchmarking and chain linking", Journal of Official Statistics, Vol. 31, pp. 627-647.

Chow, G.C. and Lin, A.I. (1971), "Best linear unbiased interpolation, distribution, and extrapolation of time series by related series", The Review of Economics and Statistics, 53(4), 372–375.

Crispino, M. and Mariani, V. (2023). "A tool to nowcast tourist overnight stays with payment data and complementary indicators", Occasional Paper, No. 746, Bank of Italy.

Di Giacinto, V., Monteforte, L., Filippone, A., Montaruli, F. and Ropele, T. (2019), "ITER: a quarterly indicator of regional economic activity in Italy", Occasional Paper, No. 489, Bank of Italy.

Emiliozzi, S., Rondinelli, C., Villa, S. (2023). "Consumption during the Covid-19 pandemic: evidence from Italian credit cards, Occasional Paper, No. 769, Bank of Italy.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). "Least angle regression." The Annals of Statistics 2004, Vol. 32, No. 2, 407–499.

Mosley, L., Eckley, I. A., and Gibberd, A. (2022a). "Sparse temporal disaggregation." Journal of the Royal Statistical Society Series A: Statistics in Society, 185(4): 2203–2233.

Mosley, L. and S. Nobari, K. (2022b). "DisaggregateTS: High-dimensional temporal disaggregation." R package version 2.0.0.

Proietti, T. (2002) "La stima rapida dei conti economici territoriali", atti della VI Conferenza Nazionale di Statistica (Roma, novembre 2002).

Schwarz, G. (1978). "Estimating the dimension of a model" Ann. Statist. 6(2): 461-464 (March, 1978). DOI: 10.1214/aos/1176344136.

van de Geer, S., Bühlmann, P.& Zhou, S. (2011) "The adaptive and the thresholded Lasso for potentially misspecified models (and a lower bound for the Lasso)." Electronic Journal of Statistics, 5, 688–749.

Zou, H. (2006) "The adaptive lasso and its oracle properties." Journal of the American Statistical Association, 101, 1418–1429.

Appendix

Concerning the model variables selection with the SFS method, we adopt the following iterative procedure:

Step 1: We estimate all the models that include the national series and only one of the identified factors; we select the model including the factor that presents the greatest significance, according to its t-test;

Step 2: We estimate all the models that include the national series, the factor selected in step 1 and only one of the remaining factors; we select the model including the factor that presents the greatest significance, according to its t-test;

Step 3: We estimate all the models that include the national series, the two factors selected in the previous steps and only one of the remaining factors; we select the model including the factor that presents the greatest significance, according to its t-test.

The alternative approach follows MEG. In order to overcome a well-known limitation of the standard Chow and Lin (1971) approach, that makes it unfeasible when the number of covariates is potentially larger than the number of yearly observations, MEG recently proposed the spTD methodology. This methodology addresses the dimensionality problem by adding a regularizing penalty on the regression coefficients β (e.g., $\ell 1$ regularizes) onto the standard GLS cost function assumed by Chow and Lin (1971) for the regression $\boldsymbol{y}_q = \boldsymbol{X}_q\beta + \boldsymbol{u}_q$, where, in MEG notation, $\boldsymbol{y}_q$ is the unobserved quarterly frequency series, $\boldsymbol{X}_q$ is the matrix of observations on the covariates and $\boldsymbol{u}_q \sim N(\boldsymbol{0}_q, \boldsymbol{V}_q)$ has the usual AR(1) representation, i.e. $u_t = \rho\, u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

The regularized estimator minimizes the following penalized objective function

$$\hat{\beta}(\lambda_n|\rho) = \arg\min_{\beta\in\mathbb{R}^d} \left\{ \left\| V_q^{-\frac{1}{2}}(y_q - X_q\beta) \right\|_2^2 + \lambda_n \|\beta\|_1 \right\} \tag{14}$$

and, differently from the standard GLS estimator, is a function of λ_n, for given autoregressive parameter ρ. So it is essential to find the most suitable value of λ_n to correctly estimate the parameters. In (14), in the estimation of $\hat{\beta}(\lambda_n|\rho)$, the solution paths of the estimator are evaluated for different values of λ_n conditioned on a fixed ρ. The solution paths are obtained using the LARS algorithm of Efron et al. (2004), as discussed in MEG.

LASSO estimators inherently exhibit a bias. To solve this problem, MEG follow Belloni and Chernozhukov (2013), by performing a refitting procedure using a least squares re-estimation that entails generating a new $n \times d^{(l)}$ sub-matrix $\boldsymbol{X'}_q$, where $d^{(l)} \le d$ from the original $n \times d$ matrix $\boldsymbol{X}_q$, with $\boldsymbol{X'}_q$, corresponding to the columns of $\boldsymbol{X}_q$, supported by $\hat{\beta}(\lambda_n^{(l)}|\rho)$, for solutions $l = 1, \dots, k$ obtained from the LARS algorithm, that produces results evaluated at a series of $\{\lambda_t\}_{t=1}^k$ points.

Finally, MEG choose the optimal estimate from $\hat{\beta}(\lambda_n^1|\rho), \ldots, \hat{\beta}(\lambda_n^k|\rho))$ using the Bayesian Information Criterion (BIC) proposed by Schwarz (1978). The motivation for suggesting this statistic over resampling methods, such as cross-validation or bootstrapping techniques, comes from the small sample size in low frequency observations. The optimal regularization is selected conditional on ρ according to

$$\hat{\lambda}(\rho) = \arg\ min_{\lambda_n(\rho)\epsilon\left\{\lambda_n^{(1)}(\rho),\ \ldots,\lambda_n^{(k)}(\rho)\right\}}\{-2\boldsymbol{\mathcal{L}}(\hat{\beta}\,(\lambda_n|\rho), \hat{\sigma}^2) + \log(n)\, K_{\lambda_n(\rho)}\} \tag{15}$$

where $K_{\lambda_n(p)} = \left|\{r: \left(\hat{\beta}(\lambda_n|\rho)_r \neq 0\}\right|$ is the degree of freedom and $\boldsymbol{\mathcal{L}}(\hat{\beta}\,(\lambda_n|\rho), \hat{\sigma}^2)$ is the log-likelihood function of the GLS regression for the β estimation in Chow and Lin (1971), which, in the presence of Gaussian errors, is given by:

$$\mathcal{L}(\hat{\beta}\,, \hat{\sigma}^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log(|\boldsymbol{S}|) - \frac{1}{2\sigma^2}(\boldsymbol{y}_q - \boldsymbol{X}_q\beta)^T\,(\boldsymbol{y}_q - \boldsymbol{X}_q\beta) \tag{16}$$

where $|\boldsymbol{S}|$ is the determinant of the Toepliz matrix **S** depending on ρ, such that $\boldsymbol{V}_q = \sigma^2\, \boldsymbol{S}$. One can then maximize this log-likelihood at the reduced-bias estimator $\hat{\beta}_p(\lambda)$ from the re-fitted LARS algorithm and using an estimator of the error variance $\hat{\sigma}^2$. By means of an iterative algorithm (see "DisaggregateTS" R package for details), MEG propose a "spTD" estimation of parameters.

In order to address the problem of multi-collinearity, MEG propose an "**adaptive-spTD**" estimation of parameters by means of another algorithm. Using the aggregated indicator matrix $\tilde{X}$ that has been GLS rotated utilizing the $\hat{\rho}$ obtained from the initial fit, they re-scale this to get $\boldsymbol{X}_{new} = \widetilde{\boldsymbol{X}}|\hat{\beta}_{\hat{\rho}}|$. A second stage regression is then performed using LARS (with BIC tuning) on (X_{new}, y) to get an estimate of $\hat{\beta}^*$. Finally, the authors scale back by $\hat{\beta}_{adapt} = \hat{\beta}^*|\hat{\beta}_{\hat{\rho}}|$ to obtain the adaptive LASSO solution. In non-GLS settings, this has been shown (van de Geer et al., 2011; Zou, 2006) to have variable selection oracle properties even when the data violates the irrepresentability condition.

TABLES AND FIGURES

Table 1. Unit root tests

Regions	Income				Consumption			
	Levels		Rates		Levels		Rates	
	ADF stat. (1)	*p*-value	ADF stat. (2)	*p*-value	ADF stat. (1)	*p*-value	ADF stat. (2)	*p*-value
Piemonte	-1,982	0,611	-2,704	0,006	-2,670	0,249	-4,528	0,000
Valle d'Aosta	-1,956	0,625	-3,146	0,002	-3,124	0,101	-4,174	0,000
Liguria	-2,109	0,541	-2,805	0,005	-3,104	0,105	-3,934	0,000
Lombardia	-2,014	0,594	-2,872	0,004	-1,982	0,611	-4,695	0,000
Prov. Aut. Bolzano	-2,482	0,337	-3,143	0,002	-1,625	0,783	-5,091	0,000
Prov. Aut. Trento	-2,456	0,350	-3,084	0,003	-2,084	0,555	-5,054	0,000
Veneto	-2,042	0,578	-3,077	0,003	-2,065	0,566	-4,229	0,000
Friuli-Venezia Giulia	-1,890	0,660	-3,130	0,002	-2,576	0,291	-3,947	0,000
Emilia-Romagna	-2,244	0,465	-2,890	0,004	-2,600	0,280	-4,013	0,000
Toscana	-2,063	0,566	-2,791	0,005	-1,975	0,615	-4,393	0,000
Umbria	-1,857	0,677	-2,800	0,005	-2,169	0,507	-3,809	0,000
Marche	-1,737	0,735	-2,861	0,005	-2,673	0,247	-3,978	0,000
Lazio	-1,928	0,640	-2,647	0,007	-2,448	0,354	-4,367	0,000
Abruzzo	-2,070	0,563	-3,039	0,003	-2,220	0,479	-3,652	0,001
Molise	-1,820	0,695	-2,673	0,007	-2,024	0,589	-3,784	0,001
Campania	-2,470	0,343	-2,200	0,019	-2,540	0,308	-3,653	0,001
Puglia	-2,128	0,531	-2,814	0,005	-2,607	0,277	-3,486	0,001
Basilicata	-2,217	0,480	-2,459	0,011	-2,752	0,215	-3,730	0,001
Calabria	-3,029	0,124	-1,599	0,062	-2,751	0,215	-3,462	0,001
Sicilia	-2,395	0,382	-1,980	0,030	-2,589	0,285	-3,134	0,002
Sardegna	-1,901	0,654	-2,790	0,005	-2,827	0,187	-3,365	0,001
Italia	-2,150	0,518	-2,497	0,010	-2,508	0,324	-4,181	0,000

(1) ADF regression with trend; lag order=2. (2) ADF regression with drift; lag order=1.

Note: This table reports the unit root (augmented Dickey-Fuller) tests for each region income (consumption) yearly time series, covering the period 1995-2022.

Table 2. Correlation and cointegration of regional household income and consumption with the corresponding national aggregate

Regions	Income				Consumption			
	Correlation		Cointegration		Correlation		Cointegration	
	Levels	Rates	Engle-Granger test (1)		Levels	Rates	Engle-Granger test (1)	
Piemonte	0.849	0.923	-2.56		0.961	0.979	-3.14	
Valle d'Aosta	0.702	0.871	-2.80		0.963	0.958	-1.64	
Liguria	0.666	0.901	-3.27	**	0.561	0.940	-1.63	
Lombardia	0.864	0.936	-2.61		0.816	0.974	-2.10	
Prov. Aut. Bolzano	0.115	0.706	-2.69		0.833	0.941	-2.65	
Prov. Aut. Trento	0.701	0.697	-2.43		0.825	0.944	-2.57	*
Veneto	0.852	0.945	-4.35		0.930	0.986	-1.78	
Friuli-Venezia Giulia	0.858	0.956	-3.88		0.961	0.965	-4.02	
Emilia-Romagna	0.924	0.924	-1.79	*	0.930	0.982	-2.05	*
Toscana	0.928	0.931	-0.80		0.948	0.980	-0.98	
Umbria	0.826	0.890	-1.69		0.973	0.948	-3.91	
Marche	0.927	0.921	-2.44		0.901	0.983	-2.48	
Lazio	0.839	0.915	-2.84		0.924	0.952	-2.35	
Abruzzo	0.790	0.851	-2.59		0.562	0.953	-3.35	
Molise	0.794	0.797	-2.54		0.908	0.891	-2.58	
Campania	0.836	0.877	-3.57		0.517	0.943	-2.07	
Puglia	0.933	0.900	-4.54		0.435	0.971	-1.78	
Basilicata	0.868	0.647	-3.27		0.786	0.962	-1.32	
Calabria	0.928	0.838	-3.58	*	0.533	0.963	-2.73	
Sicilia	0.960	0.877	-3.01		0.629	0.961	-3.11	
Sardegna	0.638	0.733	-2.35		0.777	0.964	-3.02	

(1) * and ** denote significance at the 0.05 and 0.01 level, respectively.

Note: Statistics computed on yearly data covering the period 1995-2022.

Table 3. Goodness-of-fit statistics for different specifications of the temporal disaggregation model: Income series.

Regions	Y-Xb correlation coeff. (Levels)						Y-Xb correlation coeff. (Y-o-Y rates)						ρ -parameter					
	Only natio-nal factor	Nat. factor + 1 reg. factors	Nat. factor + 2 reg. factors	Nat. factor + 3 reg. factors	spTD	Adaptive-spTD	Only natio-nal factor	Nat. factor + 1 reg. factors	Nat. factor + 2 reg. factors	Nat. factor + 3 reg. factors	spTD	Adaptive-spTD	Only natio-nal factor	Nat. factor + 1 reg. factors	Nat. factor + 2 reg. factors	Nat. factor + 3 reg. factors	spTD	Adaptive-spTD
Piemonte	0.737	0.712	0.800	0.748	0.998	0.998	0.924	0.928	0.939	0.949	0.986	0.985	0.990	0.990	0.990	0.990	0.000	0.000
Valle d'Aosta	0.595	0.822	0.920	0.925	0.997	0.997	0.858	0.933	0.952	0.952	0.976	0.976	0.990	0.985	0.971	0.966	0.404	0.404
Liguria	0.529	0.989	0.990	0.991	0.996	0.996	0.881	0.896	0.900	0.924	0.952	0.952	0.990	0.461	0.456	0.546	0.268	0.268
Lombardia	0.898	0.966	0.972	0.973	0.990	0.990	0.899	0.869	0.906	0.913	0.954	0.954	0.924	0.508	0.622	0.631	0.585	0.585
Pr. di Bolzano	0.272	0.951	0.941	0.977	0.992	0.992	0.706	0.672	0.753	0.832	0.928	0.928	0.990	0.834	0.914	0.825	0.655	0.655
Pr. di Trento	0.809	0.909	0.918	0.926	0.916	0.916	0.731	0.715	0.706	0.696	0.769	0.769	0.848	0.659	0.593	0.522	0.533	0.533
Veneto	0.899	0.983	0.984	0.984	0.999	0.999	0.954	0.961	0.963	0.969	0.994	0.993	0.966	0.744	0.754	0.792	0.000	0.000
Friuli-V. Giulia	0.747	0.972	0.985	0.986	0.990	0.988	0.916	0.903	0.924	0.945	0.938	0.934	0.981	0.437	0.050	0.527	0.000	0.000
Emilia Romagna	0.847	0.952	0.946	0.946	0.997	0.996	0.936	0.931	0.939	0.946	0.988	0.985	0.981	0.896	0.929	0.938	0.000	0.000
Toscana	0.859	0.969	0.973	0.971	0.998	0.997	0.934	0.909	0.949	0.953	0.989	0.985	0.976	0.763	0.848	0.862	0.247	0.247
Umbria	0.781	0.973	0.987	0.990	0.994	0.992	0.906	0.893	0.910	0.920	0.954	0.953	0.985	0.721	0.050	0.050	0.649	0.649
Marche	0.958	0.971	0.978	0.983	0.998	0.998	0.938	0.939	0.947	0.947	0.989	0.990	0.877	0.820	0.787	0.688	0.276	0.276
Lazio	0.891	0.948	0.937	0.947	0.998	0.997	0.884	0.895	0.906	0.918	0.984	0.971	0.943	0.914	0.933	0.938	0.264	0.264
Abruzzo	0.710	0.955	0.975	0.981	0.983	0.983	0.849	0.852	0.906	0.907	0.915	0.915	0.957	0.503	0.050	0.050	0.000	0.000
Molise	0.703	0.975	0.990	0.992	0.996	0.995	0.827	0.861	0.891	0.915	0.953	0.945	0.976	0.702	0.050	0.050	0.575	0.575
Campania	0.753	0.969	0.982	0.992	0.996	0.995	0.862	0.894	0.951	0.954	0.952	0.952	0.976	0.872	0.914	0.555	0.335	0.335
Puglia	0.940	0.971	0.969	0.972	0.996	0.995	0.922	0.935	0.942	0.951	0.979	0.973	0.905	0.829	0.867	0.881	0.000	0.000
Basilicata	0.805	0.951	0.952	0.965	0.987	0.986	0.704	0.739	0.760	0.798	0.890	0.882	0.952	0.650	0.749	0.655	0.324	0.324
Calabria	0.868	0.979	0.991	0.992	0.997	0.996	0.861	0.883	0.927	0.927	0.961	0.958	0.966	0.659	0.064	0.050	0.000	0.000
Sicilia	0.959	0.977	0.970	0.986	0.973	0.973	0.893	0.913	0.936	0.938	0.912	0.912	0.905	0.768	0.905	0.768	0.758	0.758
Sardegna	0.786	0.961	0.972	0.972	0.991	0.990	0.813	0.756	0.843	0.860	0.920	0.914	0.985	0.707	0.725	0.792	0.199	0.199

Note: In the table, Y denotes the outcome of the temporal disaggregation (income or consumption); Xb represents the fitted value of the linear regression of Y on the set of high frequency predictors X. ρ is the model residuals autoregressive coefficient, as defined in Section 2.1.

Table 4. Goodness-of-fit statistics for different specifications of the temporal disaggregation model: Consumption series.

Regions	Y-Xb correlation coeff. (Levels)						Y-Xb correlation coeff. (Y-o-Y rates)						ρ -parameter					
	Only natio-nal factor	Nat. factor + 1 reg. factors	Nat. factor + 2 reg. factors	Nat. factor + 3 reg. factors	spTD	Adaptive-spTD	Only natio-nal factor	Nat. factor + 1 reg. factors	Nat. factor + 2 reg. factors	Nat. factor + 3 reg. factors	spTD	Adaptive-spTD	Only natio-nal factor	Nat. factor + 1 reg. factors	Nat. factor + 2 reg. factors	Nat. factor + 3 reg. factors	spTD	Adaptive-spTD
Piemonte	0.965	0.994	0.996	0.996	0.999	0.998	0.985	0.983	0.986	0.986	0.995	0.993	0.957	0.489	0.442	0.447	0.697	0.697
Valle d'Aosta	0.966	0.983	0.990	0.990	0.997	0.996	0.969	0.967	0.965	0.966	0.988	0.984	0.896	0.702	0.050	0.050	0.000	0.000
Liguria	0.590	0.976	0.984	0.986	0.994	0.994	0.953	0.953	0.953	0.954	0.984	0.984	0.985	0.655	0.277	0.050	0.694	0.694
Lombardia	0.826	0.970	0.994	0.996	0.999	0.999	0.982	0.981	0.979	0.985	0.989	0.989	0.990	0.957	0.744	0.664	0.000	0.000
Pr. di Bolzano	0.847	0.865	0.969	0.978	0.997	0.996	0.960	0.983	0.992	0.994	0.993	0.992	0.976	0.976	0.957	0.943	0.000	0.000
Pr. di Trento	0.840	0.982	0.981	0.990	0.997	0.997	0.962	0.974	0.985	0.987	0.993	0.992	0.971	0.725	0.877	0.744	0.600	0.600
Veneto	0.933	0.992	0.996	0.997	1.000	1.000	0.990	0.986	0.987	0.991	0.999	0.998	0.990	0.867	0.688	0.730	0.000	0.000
Friuli-V. Giulia	0.963	0.976	0.987	0.989	0.999	0.999	0.975	0.978	0.980	0.980	0.996	0.996	0.947	0.924	0.867	0.848	0.000	0.000
Emilia Romagna	0.932	0.966	0.987	0.992	0.999	0.999	0.986	0.988	0.988	0.991	0.995	0.995	0.990	0.990	0.990	0.985	0.523	0.523
Toscana	0.951	0.996	0.997	0.997	0.997	0.997	0.985	0.984	0.986	0.990	0.986	0.986	0.971	0.409	0.215	0.404	0.022	0.022
Umbria	0.974	0.985	0.988	0.990	0.998	0.998	0.963	0.972	0.974	0.977	0.987	0.988	0.900	0.834	0.801	0.796	0.000	0.000
Marche	0.908	0.992	0.979	0.994	0.998	0.998	0.988	0.987	0.989	0.990	0.991	0.991	0.985	0.782	0.943	0.815	0.000	0.000
Lazio	0.934	0.976	0.982	0.989	0.998	0.998	0.965	0.968	0.975	0.980	0.990	0.990	0.957	0.881	0.881	0.825	0.000	0.000
Abruzzo	0.596	0.978	0.983	0.980	0.997	0.997	0.967	0.975	0.976	0.977	0.993	0.993	0.990	0.900	0.877	0.905	0.742	0.742
Molise	0.911	0.953	0.976	0.983	0.997	0.995	0.920	0.944	0.944	0.954	0.990	0.981	0.914	0.877	0.659	0.560	0.543	0.543
Campania	0.545	0.970	0.978	0.983	0.999	0.999	0.959	0.951	0.972	0.978	0.991	0.990	0.990	0.910	0.919	0.910	0.000	0.000
Puglia	0.464	0.978	0.980	0.973	1.000	1.000	0.979	0.982	0.985	0.986	0.998	0.998	0.990	0.966	0.976	0.985	0.491	0.491
Basilicata	0.801	0.973	0.970	0.976	0.998	0.996	0.973	0.973	0.978	0.979	0.991	0.989	0.985	0.877	0.929	0.910	0.000	0.000
Calabria	0.567	0.957	0.994	0.994	0.999	0.999	0.973	0.971	0.976	0.975	0.993	0.993	0.990	0.971	0.050	0.050	0.569	0.569
Sicilia	0.653	0.853	0.948	0.978	0.999	0.998	0.972	0.975	0.974	0.977	0.990	0.988	0.990	0.990	0.981	0.966	0.007	0.007
Sardegna	0.796	0.981	0.962	0.992	0.999	0.998	0.974	0.966	0.973	0.969	0.992	0.988	0.985	0.825	0.943	0.541	0.388	0.388

Note: In the table, Y denotes the outcome of the temporal disaggregation (income or consumption); Xb represents the fitted value of the linear regression of Y on the set of high frequency predictors X. ρ is the model residuals autoregressive coefficient, as defined in Section 2.1.

Table 5. Mean absolute error (MAE) of the yearly rate of growth forecast: household income series. (1)

Regions	Unbalanced forecasts				Balanced forecasts			
	SFS (2) - two local factors	SFS (2) – three local factors	spTD	Adaptive-spTD	SFS (2) - two local factors	SFS (2) – three local factors	spTD	Adaptive-spTD
Piemonte	1.05	1.33	1.16	1.29	0.49	0.56	0.58	0.60
Valle d'Aosta	1.12	1.28	1.71	1.77	0.94	0.81	0.98	0.99
Liguria	0.61	0.91	0.98	0.81	0.69	0.70	0.69	0.57
Lombardia	0.50	0.66	0.72	0.49	0.43	0.43	0.56	0.52
Prov. Aut. Bolzano	2.11	1.87	2.30	2.20	1.29	1.12	1.39	1.39
Prov. Aut. Trento	0.70	0.90	0.46	0.74	0.40	0.48	0.43	0.58
Veneto	0.61	0.82	1.21	1.11	0.48	0.55	0.84	0.79
Friuli-Venezia Giulia	0.60	0.81	0.73	0.81	0.51	0.67	0.59	0.60
Emilia-Romagna	0.48	0.51	0.49	0.62	0.34	0.36	0.46	0.49
Toscana	0.98	0.87	0.89	0.84	0.74	0.83	0.89	0.85
Umbria	1.36	1.64	1.86	1.94	1.01	1.15	1.25	1.26
Marche	0.94	1.10	0.80	0.76	0.55	0.62	0.58	0.54
Lazio	0.56	0.56	1.20	1.28	0.26	0.28	0.56	0.57
Abruzzo	1.11	1.16	1.60	1.49	0.76	0.80	0.98	0.97
Molise	2.27	2.24	1.75	1.66	1.39	1.36	1.11	1.12
Campania	1.19	1.17	1.55	1.87	0.63	0.62	0.74	0.84
Puglia	0.85	0.98	0.76	0.75	0.64	0.65	0.52	0.48
Basilicata	1.37	1.26	1.90	1.73	1.23	1.32	1.40	1.38
Calabria	1.22	0.99	0.91	0.84	0.67	0.54	0.60	0.64
Sicilia	0.84	0.76	1.50	1.42	0.67	0.62	0.82	0.82
Sardegna	1.20	1.29	1.53	1.58	1.17	1.19	1.16	1.20
Average across regions	1.03	1.10	1.24	1.24	0.73	0.75	0.82	0.82

(1) The forecast exercise spans a 7-years period, from 2016 to 2022. – (2) SFS stands for Stepwise Forward Selection.

Note: The balanced forecast is computed from the unbalanced forecast, by imposing the constraint that the regional series sum to the national (known) aggregate.

Table 6. Mean absolute error (MAE) of the yearly rate of growth forecast: household consumption series. (1)

Regions	Unbalanced forecasts				Balanced forecasts			
	SFS (2) - two local factors	SFS (2) – three local factors	spTD	adaptive-spTD	SFS (2) - two local factors	SFS (2) – three local factors	spTD	adaptive-spTD
Piemonte	0.57	0.63	1.16	1.29	0.39	0.36	0.58	0.60
Valle d'Aosta	1.20	1.65	1.71	1.77	0.75	0.77	0.98	0.99
Liguria	0.99	0.91	0.98	0.81	0.50	0.42	0.69	0.57
Lombardia	0.72	0.50	0.72	0.49	0.41	0.39	0.56	0.52
Prov. Aut. Bolzano	2.07	2.31	2.30	2.20	1.45	1.55	1.39	1.39
Prov. Aut. Trento	1.63	1.47	0.46	0.74	1.15	1.02	0.43	0.58
Veneto	0.54	0.39	1.21	1.11	0.46	0.38	0.84	0.79
Friuli-Venezia Giulia	0.93	1.12	0.73	0.81	0.45	0.56	0.59	0.60
Emilia-Romagna	0.51	0.69	0.49	0.62	0.20	0.24	0.46	0.49
Toscana	0.57	0.26	0.89	0.84	0.42	0.34	0.89	0.85
Umbria	0.59	0.73	1.86	1.94	0.45	0.43	1.25	1.26
Marche	1.04	0.80	0.80	0.76	0.52	0.43	0.58	0.54
Lazio	2.36	2.22	1.20	1.28	1.22	1.17	0.56	0.57
Abruzzo	0.71	0.62	1.60	1.49	0.35	0.23	0.98	0.97
Molise	1.08	0.93	1.75	1.66	0.58	0.53	1.11	1.12
Campania	1.25	0.99	1.55	1.87	0.57	0.51	0.74	0.84
Puglia	1.65	1.76	0.76	0.75	0.84	0.96	0.52	0.48
Basilicata	0.69	0.80	1.90	1.73	0.46	0.52	1.40	1.38
Calabria	2.19	2.29	0.91	0.84	1.03	1.04	0.60	0.64
Sicilia	1.82	1.84	1.50	1.42	0.92	0.96	0.82	0.82
Sardegna	1.36	1.37	1.53	1.58	0.37	0.43	1.16	1.20
Average across regions	1.16	1.15	1.24	1.24	0.64	0.63	0.82	0.82

(1) The forecast exercise spans a 7-years period, from 2016 to 2022. – (2) SFS stands for Stepwise Forward Selection.

Note: The balanced forecast is computed from the unbalanced forecast, by imposing the constraint that the regional series sum to the national (known) aggregate.

Figure 1. Household income time series by region and for Italy (observed annual data).

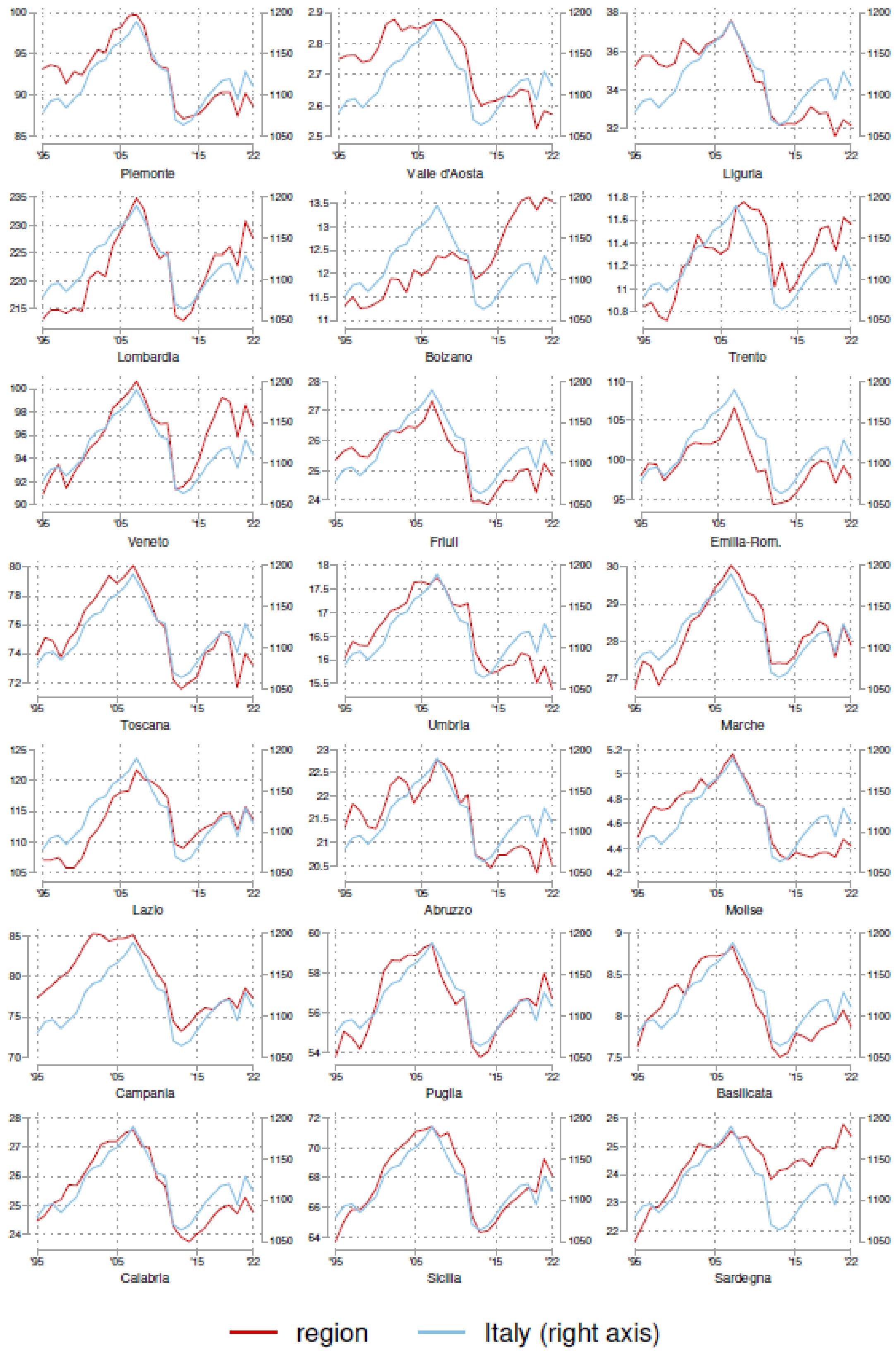


Figure 2. Household consumption time series by region and for Italy (observed annual data).

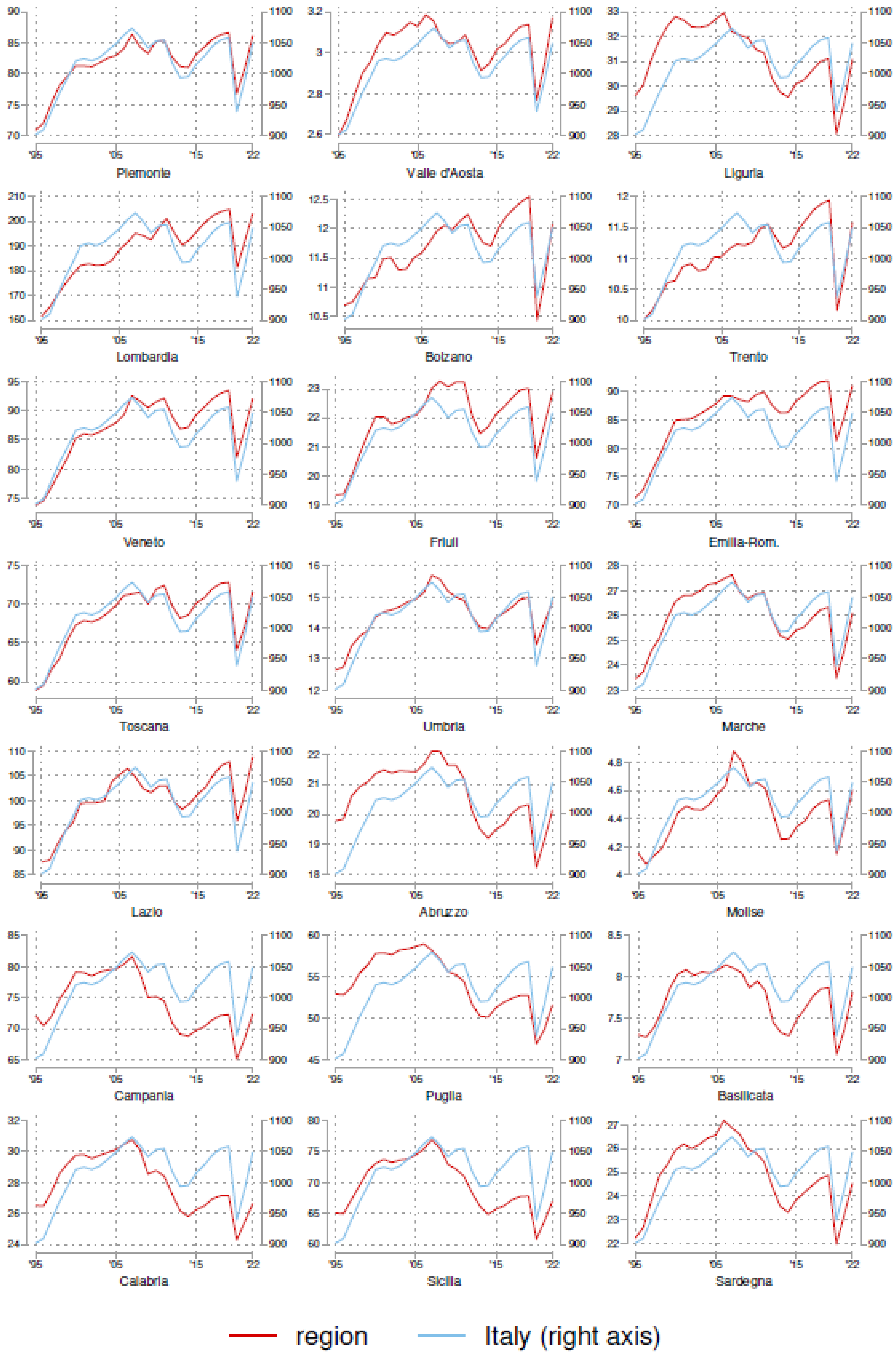


Figure 3 – First three common factors for each region (income)

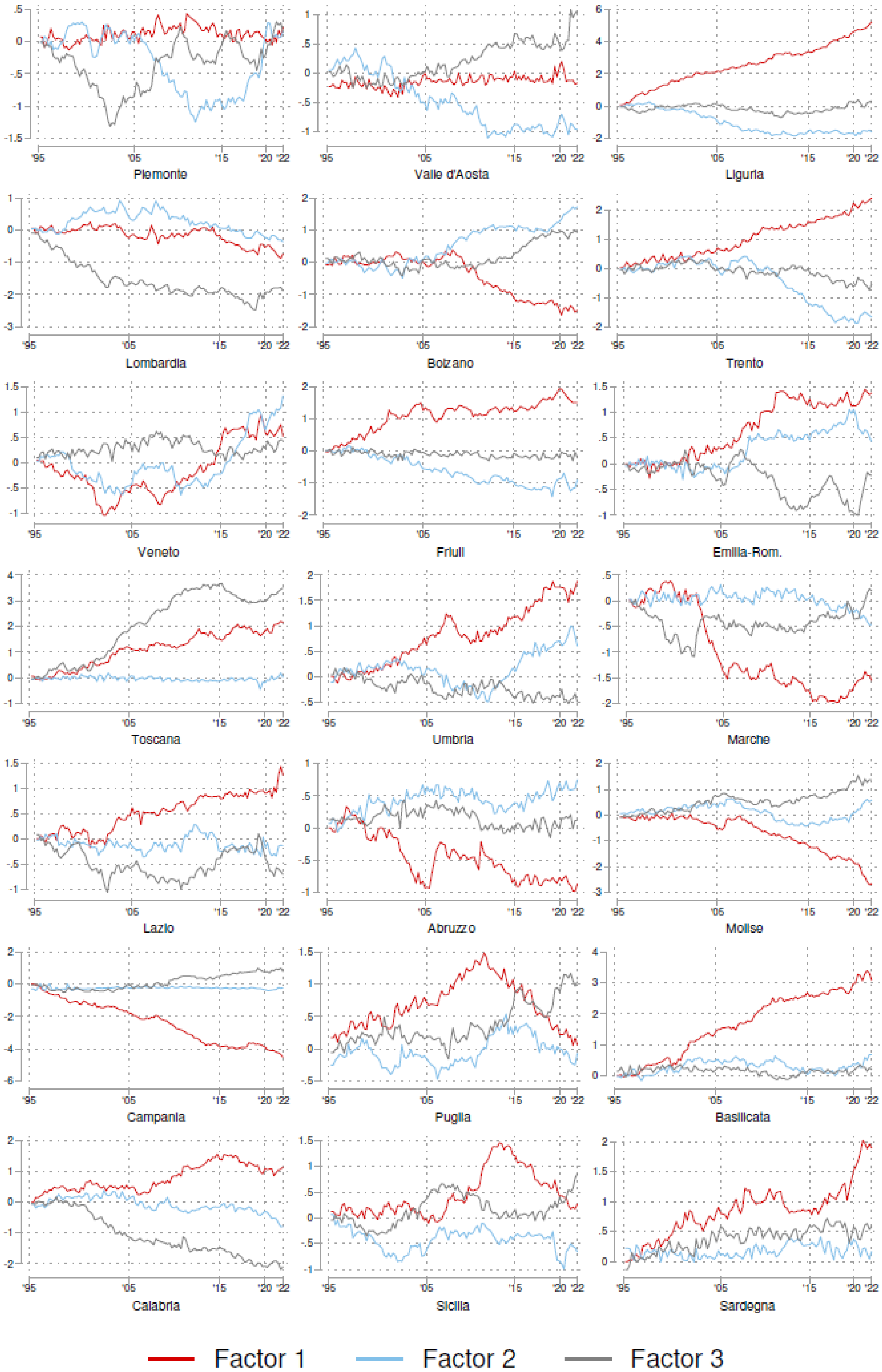


Figure 4 – First three common factors for each region (consumption)

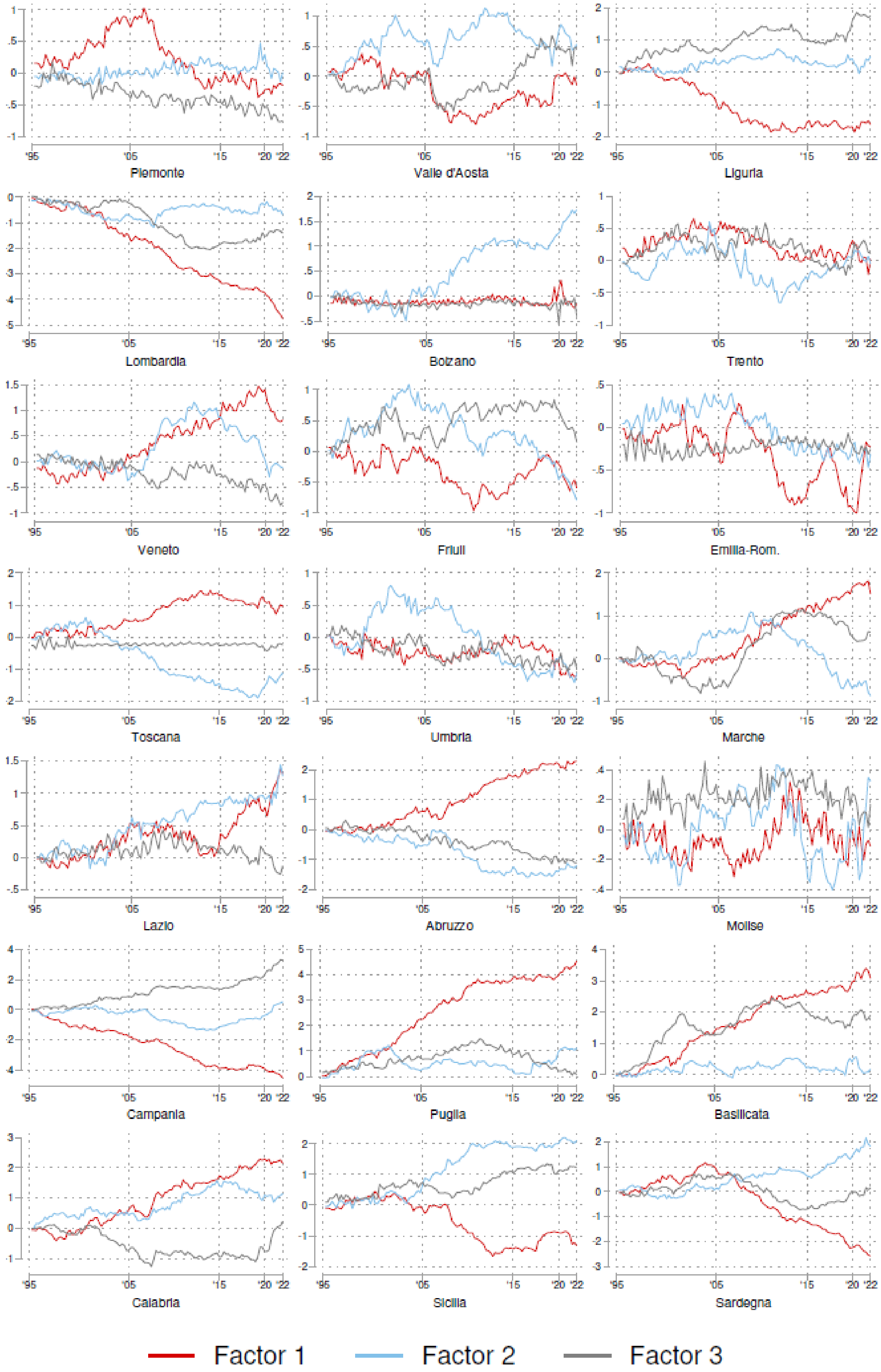
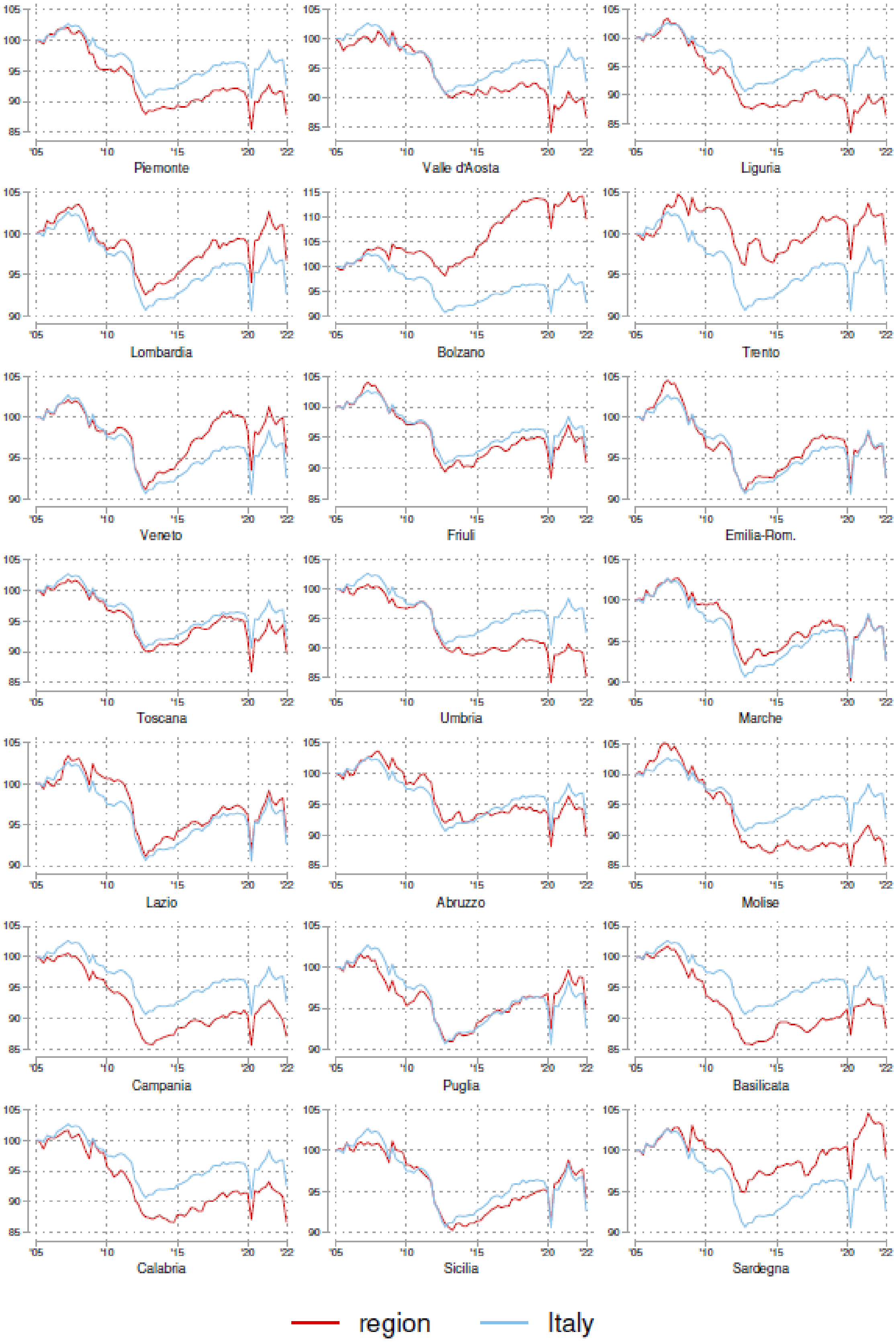
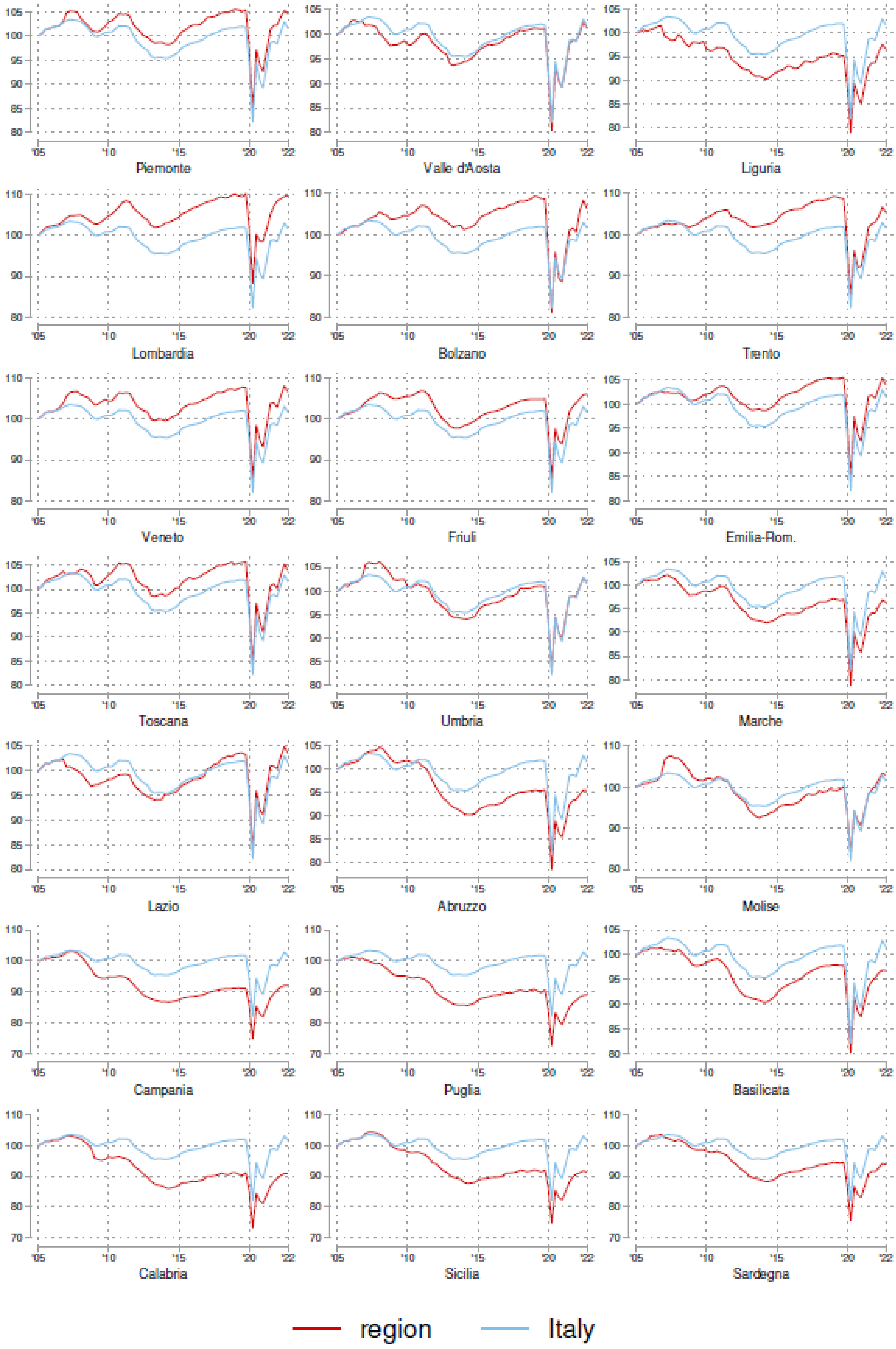


Figure 5 – Quarterly regional indicators for disposable income



Note: The data for Italy are observed (official national statistics), while the data for each region are estimated.

Figure 6 – Quarterly regional indicators for consumption



Note: The data for Italy are observed (official national statistics), while the data for each region are estimated.

Figure 7 – Contrasting ITER-red and official statistics year by year

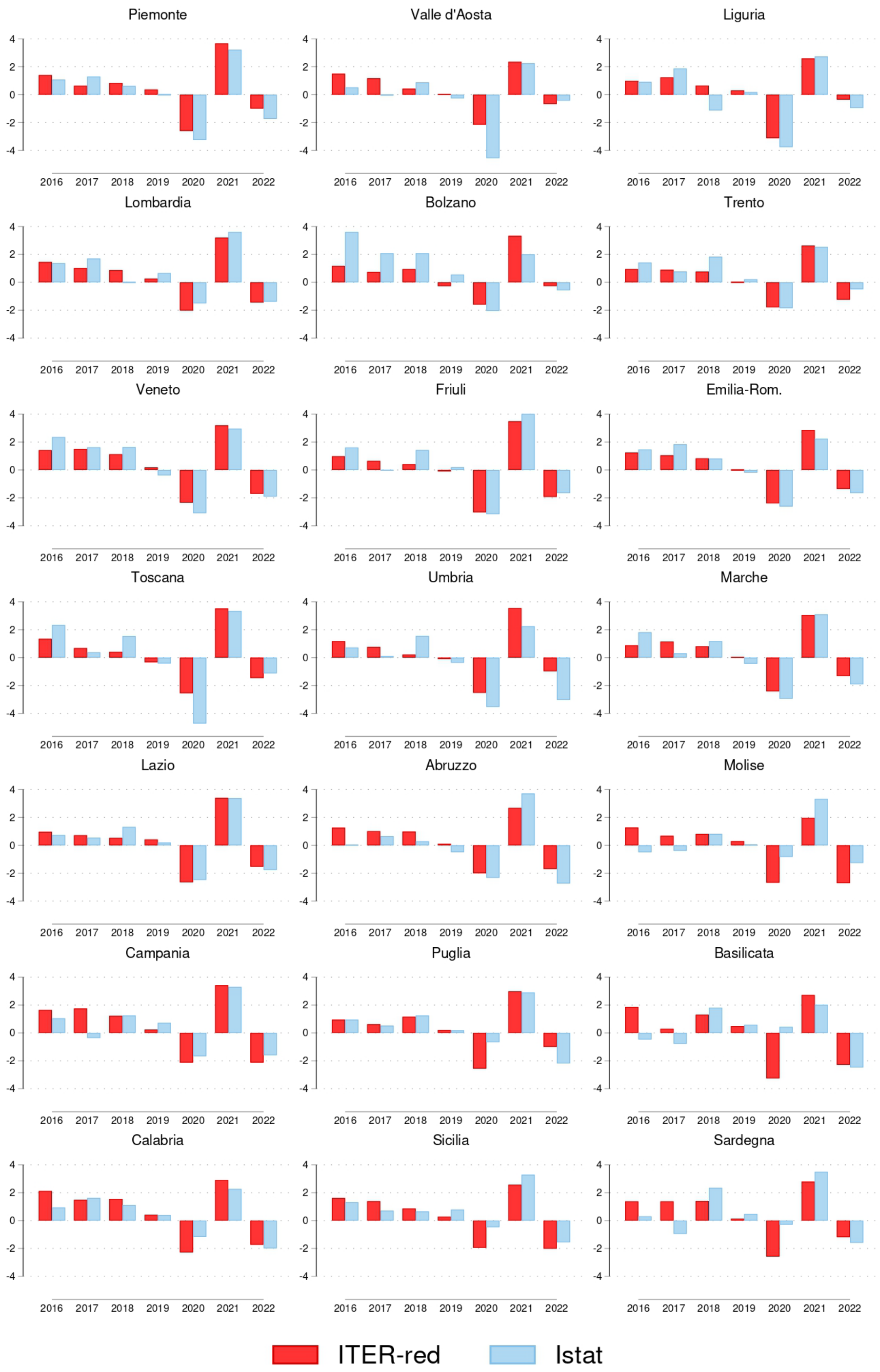


Figure 8 – Contrasting ITER-con and official statistics year by year