

Article

A Set of State–Space Models at a High Disaggregation Level to Forecast Italian Industrial Production

Riccardo Corradini

DIPS Department, Division for Data Analysis and Economic, Social and Environmental Research, ISTAT Italian National Institute of Statistics, 00198 Rome, Italy; corradin@istat.it

Received: 11 June 2019; Accepted: 11 November 2019; Published: 18 November 2019



Abstract: Normally, econometric models that forecast the Italian Industrial Production Index do not exploit information already available at time $t + 1$ for their own main industry groupings. The new strategy proposed here uses state–space models and aggregates the estimates to obtain improved results. The performance of disaggregated models is compared at the same time with a popular benchmark model, a univariate model tailored on the whole index, with persistent not formally registered holidays, a vector autoregressive moving average model exploiting all information published on the web for main industry groupings. Tests for superior predictive ability confirm the supremacy of the aggregated forecasts over three steps horizon using absolute forecast error and quadratic forecast error as a loss function. The datasets are available online.

Keywords: Industrial Production Index; forecasting; disaggregation; Kalman filter

1. Introduction

Forecasting industrial production can be a difficult task, but forecasting the sub-components of industrial production at a high disaggregation level can be even more challenging for researchers. This happens because there is some information available at the disaggregated level, and unless it has movements similar to the sub-components, there is the risk of worsening forecast results by simply exploiting the past. Gross data on the Italian Industrial Production Index at a higher disaggregated level are analyzed, exploiting the high correlation with other times series when they are available. Low forecasting performance on average is found when poor or no information is available, consistent with existing literature, but larger improvements were observed for the disaggregated components that face a richer correlation with other variables on their past and on the one-step-ahead prediction.

The related literature (see Bulligan et al. [1]) shows two main sets of models. The first set exploits quantitative data when they are available for the whole index or for its disaggregated components, using regression methods (see, for instance, Marchetti et al. [2] and Bodo et al. [3]) or seemingly unrelated equations methods as in the work of Bruno et al. [4]. The second set exploits so-called common factors that summarize a big set of survey data about industrial production as predictors.

Costantini [5] showed how “*The state-space model yields superior forecasts among the factor models*”.

To the best of my knowledge, this is the only study that compared the short-term forecasting performance of a naive autoregressive model applied to the Italian Industrial Production Index with a set of state-space models applied to its own main industry groupings.

The effort is motivated to understand the short-term evolution of the Italian economy. It is mandatory to continuously review the estimates to monitor the evolution of quarterly estimates of gross domestic product. There is a trade-off between the need to provide industrial production forecasts closely related with the future evolution of gross domestic product in a short time and doing such a job by exploiting all the information available, as much as possible, with the most suitable econometric tools.

Brunes-Lesage et al. [6] pointed out, when analyzing the French Industrial Production Index, that *"The IIP, however, is characterized by a significant publication delay, around 40 days after the end of the reference month for the main European countries, and the first IIP estimations are often revised significantly. Thus, it is less useful for short-term forecasting exercises."* Italian data are not so prone to high revisions, but they are also characterized by the same delay (40 days). The aim of this paper is, therefore, to propose several models designed to forecast the current-month Italian Industrial Production Index (nowcast) using the data listed in Table 1 and to forecast the Italian Industrial Production Index for the next two months after the present time month.

It is shown in further detail using a recursive forecast window starting from January 2001 and moving from January 2008 up to December 2018 that a popular autoregressive of order three over the log seasonal differences benchmark used by Bulligan et al. [1] can be beaten at any disaggregated level. (The Destatis Truck Toll Mileage Index starts from January 2005 and the time series for natural gas from January 2006. Nevertheless, for 2005, we use for natural gas the level given by Italian Ministry of Economic Development. For 2004, we give data about natural gas transported the same monthly percentage of industrial gas registered for the year 2005. For 2001–2004, a naive backcast is carried out for the Destatis Truck Toll Mileage Index using seasonal growth rates of published German Industrial Index of Intermediate Goods. The same strategy is used to backcast the level of production of cars from 2001 to 2007 using consolidated published data from Destatis about production of cars.)

It is shown how aggregating the estimates can obtain more competitive results than the above-mentioned autoregressive model over a three step horizon. We exploit state-space models, focusing mainly on the Truck Toll Mileage Index in Germany (see Askita et al. [7] and Cox et al. [8]) to forecast the Industrial Index of Intermediate Goods and of Capital Goods. Non-durable goods exploit the industrial gas data. Durable goods forecasts exploit the Destatis Truck Toll Mileage Index. Conversely, for the production of electricity, gas, steam and air conditioning supply, due to abundance of daily and monthly Italian data, a set of ad-hoc models at different frequencies is used.

The paper contributes to the research community for the following reasons:

- A workhouse is beaten by the aggregated model. However, it is still competitive at the third step.
- A model tailored on Italian holidays history is beaten by the aggregated model.
- All data are freely available on the web without any restrictions.
- It uses open source software matrix programming language GNU Octave (see <https://www.gnu.org/software/octave/>) to enforce transparency of results and to elicit debate among researchers all over the world.
- For the period before 2015, where it is impossible to use the same fixed weights, a solution is proposed.
- Computational algorithms and state-space models have never been applied in such a disaggregated context for the Italian Production Index and are not available in standard commercial packages.
- It shows how it is possible to cast into a state-space model (without imposing restrictions) a vector autoregressive moving average applied to the whole index that exploits as endogenous variables all freely available data at $t + 1$ on the web concerning Main Industry Groupings disaggregation from January 2001 to December 2018.
- For the first time, model confidence set procedure of Hansen et al. (see [9]) is applied to Italian Industrial Production Index as well as its sub-components to check the superior predictive ability of aggregated forecasts over a class of competing econometric models.

The remainder of the paper is organized as follows. In Section 2, the data used are presented with the weighted structure of the Italian Production Index. In Section 3, the barebone model and the barebone model enhanced by stochastic regressors are presented. The last one is used for an ad-hoc forecasting study concerning intermediate goods, capital goods and non-durable goods. In Section 4, more recent models are upgraded inserting most of the data freely available on the web used in the disaggregated models. In Section 5, the basic assumptions of the models are inspected. In Section 6,

the Kalman filter and the smoothing algorithm are summarized. Section 7 points out the main results using the predictive ability tests (see Section 3.7). Finally, some conclusions peculiar to the Italian Industrial Production Index are drawn.

2. Data

Figure 1 plots the first log seasonal difference of the following indexes: Industrial Production Index for Intermediate Goods, Industrial Production Index for Capital goods and the German Truck Toll Mileage Index. The last variable is always available at time $t + 1$. As shown in Figure 2, intermediate goods and capital goods account for 61 per cent of the whole index. Ad-hoc bivariate seemingly unrelated time series equations are used (see example about car drivers' accidents in the work of Durbin et al. [10] and Section 3.1). The assumption is that these goods are transported in these countries (Italy and Germany) monthly and that there is a relationship among their seasonal growth rates. A similar strategy is considered for the Industrial Production Index of Durable Goods. Bivariate seemingly unrelated time series equations (see Section 3.1) are composed by the German Truck Toll Mileage Index and the Industrial Production Index of Durable Goods. Similarly, for the Industrial Production Index of Non-durable Goods, the seemingly unrelated time series equations are applied to the cumulated monthly data about the industrial natural gas used. On the contrary, a lot of information is ready at the time of the publication on ISTAT's website concerning production electricity and other sub-components (Figure 3 shows the weights of sub-components). The daily data about consumption of electricity can be cumulated to obtain a preliminary estimate of the monthly data that will be released roughly 15–20 days later. Data about compressed natural gas are less prone to revision and do not differ too much from their preliminary value when cumulated to obtain their monthly value. To obtain an estimate of main industry grouping in the Industrial Production Index of Electricity, gas, steam and air conditioning supply, we must consider four sub-components:

- Extraction of crude petroleum and natural gas
- Manufacture of coke and refined petroleum products
- Electric power generation, transmission and distribution
- Manufacture of gas; distribution of gaseous fuels through mains.

Most of the time, extraction of crude petroleum happens together with natural gas. Unfortunately, no information was available at time $t + 1$ for crude petroleum, but we do have daily data about the extraction of natural gas from 2006 up to now. These data are cumulated and inserted into seemingly unrelated time series equations (see, for instance, the example about car drivers killed or seriously injured in Great Britain from January 1969 to December 1984 in [10] and Section 3.1). For $t + 2$, roughly 25 daily observations (at that time, all monthly data are ready; see Table 1) are always available and are appended to daily time series about the production of compressed natural gas. Manufacture of coke and refined petroleum products is estimated by an autoregressive moving average identification procedure described in Gómez et al. [11] applied to ISTAT's monthly data. Electric power generation, transmission and distribution is somewhat more complicated.

At time $t + 1$, production of electricity data from Terna's website and roughly 25 daily observations of consumption are available because two endogenous variables are available at time $t + 1$:

- Consumption of electricity on monthly and daily basis
- Production of electricity

A conditional vector autoregressive moving average model is used to estimate the index of electric power generation, transmission and distribution at time $t + 1$. For $t + 2$ and $t + 3$, a naive autoregressive model of order three using first seasonal weekly differences applied to daily data about consumption and thermoelectric, national natural gas production and transportation allows us to have more reliable estimates for production of electricity, extraction of petroleum and natural gas and finally distribution of gaseous fuels through mains at time $t + 2$ and $t + 3$ (see the Appendixes A and B for further details).

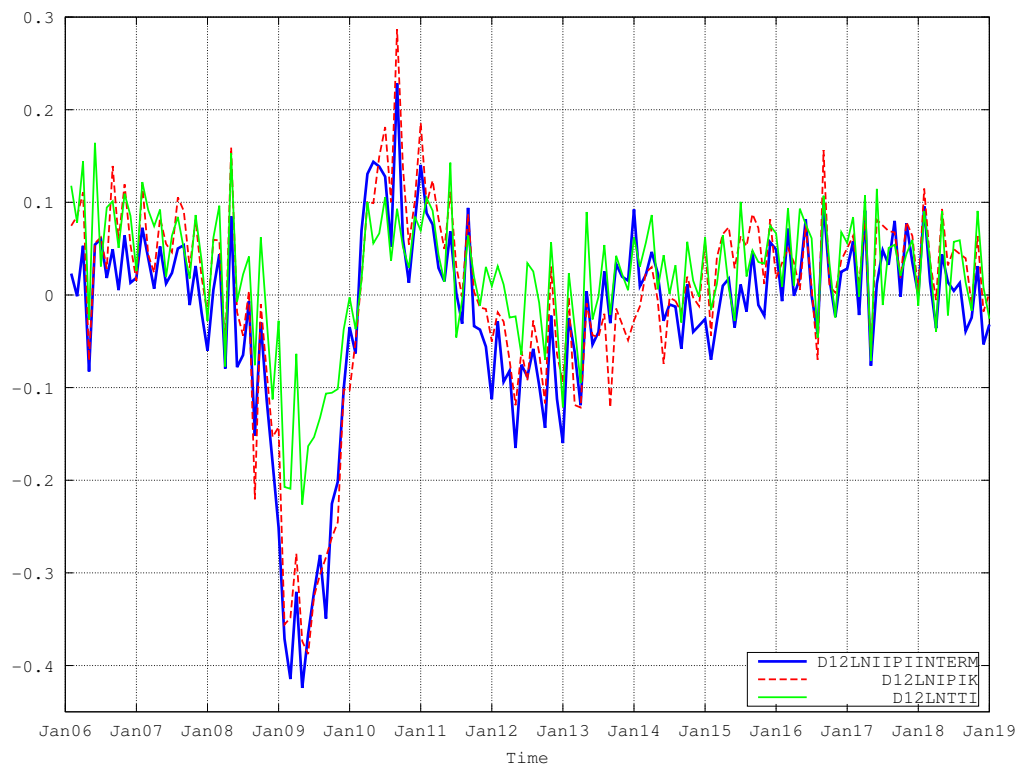


Figure 1. This graph plots first log seasonal differences of Industrial Production Index of Intermediate Goods, Capital Goods and Destatis Truck Toll Mileage Index from 2006:01 to 2018:12.

Table 1. Information sources of data.

Name	Series	Sources	Freq.	Time ^a
IPI	5	ISTAT	Monthly	-
IPI of Electricity, gas, steam and air conditioning supply	4	ISTAT	Monthly	-
GigaWatt Electricity net production	1	TERNA	Monthly	+20+25 days
GigaWatt Electricity Consumption	1	TERNA	Monthly	+35+40 days
GigaWatt Electricity Consumption	1	TERNA	Daily	Every Day
Total m3 of Compressed Natural Gas transported	1	SNAM	Daily	Every Day
Production of m3 of Compressed Natural Gas	1	SNAM	Daily	Every Day
M3 of Compressed Natural Gas for Thermoelectric use	1	SNAM	Daily	Every Day
M3 of Compressed Natural Gas for Industrial use	1	SNAM	Daily	Every Day
Registration of Light Commercial Vehicles in Italy	1	ACI	Monthly	+35+40 days
Registration of Light Commercial Vehicles in Spain	1	ANFAC	Monthly	+35+40 days
Registration of Light Commercial Vehicles in France	1	CCFA	Monthly	+35+40 days
Registration of Light Commercial Vehicles in Germany	1	KBA	Monthly	+35+40 days
Truck Toll Mileage Index in Germany	1	DESTATIS	Monthly	+35+40 days
IPI of Manufacture of motor vehicles, trailers and semi-trailers	1	ISTAT	Monthly	-
Production of passengers cars in Germany	1	VDA	Monthly	+35+40 days

^a Publication corresponds to amounts of days before the official release on ISTAT website.

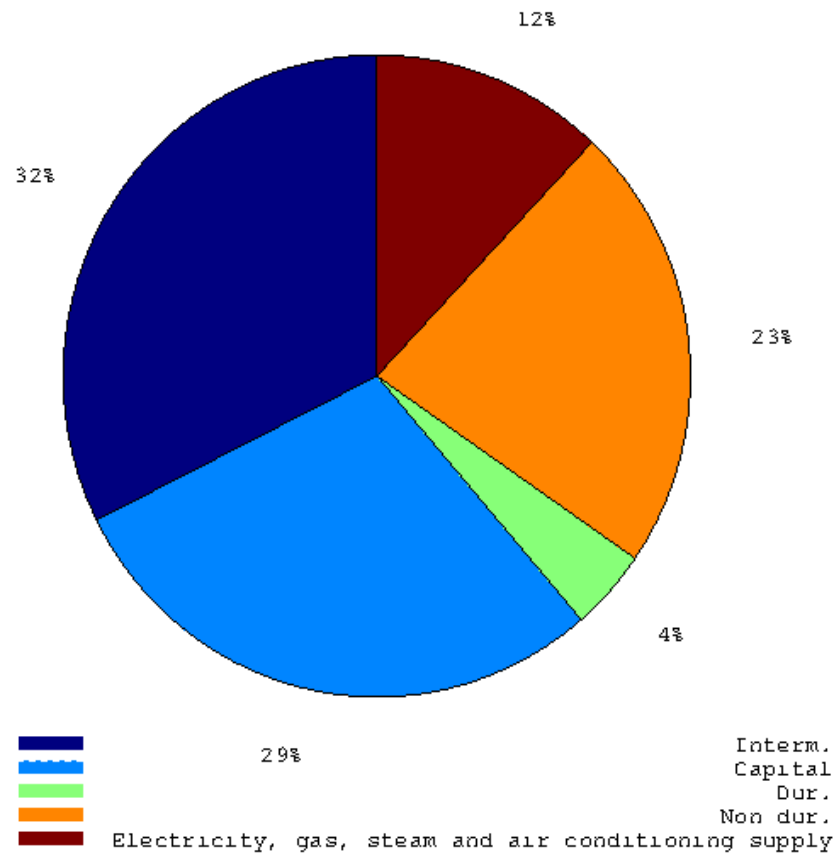


Figure 2. This pie graph shows the weights of main industry groupings in the Italian Industrial Production Index.

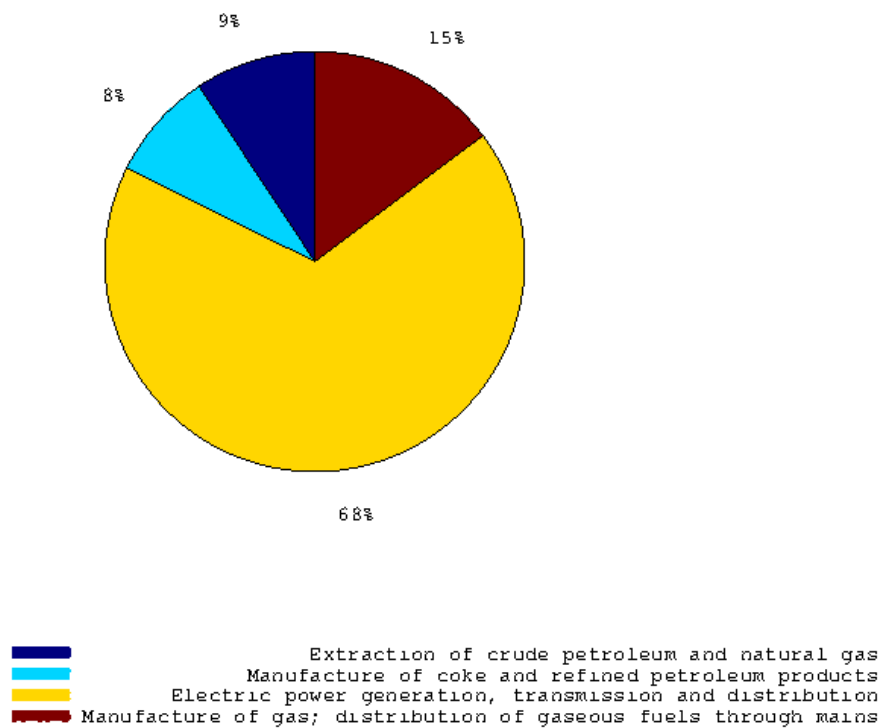


Figure 3. This pie graph shows the weights of every sub-component concerning the Italian Industrial Production Index of Electricity, gas, steam and air conditioning supply.

3. Models Used

3.1. Barebone Model

First, our model incorporates regression variables into structural models. Thus, it has the following structure.

$$Y_t = X_t\beta + w_t \quad (1)$$

where w_t takes the following general structure (as described by Gómez [12])

$$w_t = p_t + s_t + u_t + v_t + e_t \quad (2)$$

where p_t is the trend, s_t is the seasonal, u_t is the cyclical, v_t is the autoregressive and e_t is the irregular component.

The trend component can be summarized in the following way:

$$p_{t+1} = p_t + b_t + c_t \quad (3)$$

$$b_{t+1} = b_t + d_t \quad (4)$$

where p_t is the level and b_t is the slope of the trend, and c_t and d_t are two mutually and serially uncorrelated sequences of random variables with zero mean and variances σ_c^2 and σ_d^2 , respectively.

The component s_t follows trigonometric seasonality

$$s_t = \sum_{i=1}^{[f/2]} s_{i,t} \quad (5)$$

where f is amounts of observations per year. In our case, $f = 12$. Each component follows the model

$$\begin{bmatrix} s_{i,t+1} \\ s_{i,t+1}^* \end{bmatrix} = \begin{bmatrix} \cos\omega_i & \sin\omega_i \\ -\sin\omega_i & \cos\omega_i \end{bmatrix} \cdot \begin{bmatrix} s_{i,t} \\ s_{i,t}^* \end{bmatrix} + \begin{bmatrix} j_{i,t} \\ j_{i,t}^* \end{bmatrix} \quad (6)$$

It is assumed that all seasonal components have a common variance, $\sigma_i^2 = \sigma_s^2$, $i = 1, 2, \dots, [f/2]$. To simplify things, the cyclical and autoregressive components are not present. It is straightforward to extend Model 1 into a multivariate one using the Kronecker product.

Thus,

$$Y_t = X_t\beta + W_t \quad (7)$$

where X_t is a regression matrix and W_t is a multivariate structural model. These last models can be obtained easily from univariate structural models using the Kronecker product (see Chapter 8 of Harvey [13]).

Now let us think for a while X_t follows a stochastic process. More accurately, following Gómez [12],

$$x_{t+1} = Fx_t + Gu_t \quad (8)$$

$$Y_t = X_t\beta + Hx_t + Ju_t, t = 1, 2, \dots, n \quad (9)$$

where X_t is the regression matrix. Since X_t follows a stochastic process, the term $X_t\beta$ can be partitioned into two terms.

$$X_t\beta = M_t\gamma + N_t\delta = M_t\gamma + (\delta' \otimes I)\text{vec}(N_t) \quad (10)$$

It is possible to rewrite the last equation in the following way

$$X_t\beta = M_t\gamma + Ru_t \quad (11)$$

where M_t is a matrix containing the non-stochastic inputs, while u_t is the matrix containing the stochastic inputs. Finally, γ and R are the exogenous inputs regressors coefficients and the stochastic input regressors coefficients, respectively.

If we consider the transition equation and the measurement equation for the stochastic inputs u_t ,

$$x_{t+1}^u = F^u x_t^u + G^u v_t \quad (12)$$

$$u_t = X_t^u \beta^u + H^u x_t^u + J^u v_t, t = 1, 2, \dots, n \quad (13)$$

With a little bit of algebra combining the model for input with the general model, we obtain the following transition equation and measurement equation

$$\begin{bmatrix} x_{t+1} \\ x_{t+1}^u \end{bmatrix} = \begin{bmatrix} F_t & 0 \\ 0 & F_t^u \end{bmatrix} \begin{bmatrix} x_t \\ x_t^u \end{bmatrix} + \begin{bmatrix} G_t & 0 \\ 0 & G_t^u \end{bmatrix} \begin{bmatrix} u_t \\ v_t \end{bmatrix} \quad (14)$$

$$\begin{bmatrix} Y_t \\ u_t \end{bmatrix} = \begin{bmatrix} M_t & R X_t^u \\ 0 & X_t^u \end{bmatrix} \begin{bmatrix} \gamma \\ \beta^u \end{bmatrix} + \begin{bmatrix} H_t & R H_t^u \\ 0 & H_t^u \end{bmatrix} \begin{bmatrix} x_t \\ x_t^u \end{bmatrix} + \begin{bmatrix} J_t & R J_t^u \\ 0 & J_t^u \end{bmatrix} \begin{bmatrix} u_t \\ v_t \end{bmatrix} \quad (15)$$

The initial conditions for this combined model depend on the initial states x_1 and x_1^u . Since x_1 is known, we have to compute x_1^u . This can be done solving the Lyapunov equation using as input the matrices F^u and G^u . Further details are indicated in the work by Gómez [12]. To obtain initial conditions for the Kalman filter, the mean and the covariance matrix of the initial state vector x_1^u are needed. If the series is stationary, the mean is obviously zero. As for the covariance matrix, letting $\text{Var}(x_{u1}) = V$, the matrix V satisfies the Lyapunov equation

$$V = F^u \cdot V \cdot F^{u'} + G^u \cdot G^{u'}; \quad (16)$$

A more detailed explanation was given by Gómez [14] (Paragraph 4.14.2 about initial conditions in the time invariant case).

3.2. Stochastic Regressors Inside the Barebone Model

The transition and measurement equations shown in Section 3.1 take as given state-space system matrices F_t^u , G_t^u , X_t^u , H_t^u , and J_t^u . At time $t + 1$, the German Truck Toll Index is available. Unfortunately, no information is available for the Italian Industrial Production Index for the manufacture of motor vehicles, trailers and semi-trailers.

Nevertheless, we assume there is a feasible autoregressive moving average seasonally integrated model that could be estimated recursively over time by the automatic identification procedure described by Gómez et al. [11]. For production of cars in Germany, the same procedure is applied. For the sake of interest, we call this system “star system”, where the respective state-space system matrices are cast on the main diagonal to give final results of the aforementioned F_t^u , G_t^u , X_t^u , H_t^u , and J_t^u . Figure 4 shows the flowchart of the computer source code used to obtain the in-sample state-space stochastic input system.

Once the Italian and German data are loaded in memory, the automatic identification procedure stores the results about their respective state-space forms. If the algorithms crashes, it is mandatory to use the logarithmic transformation (see Gómez [15]) and the property of log normal distribution to bring those data on the original level if it is desirable. Finally, the matrices are cast on the main diagonal of F_t^u , G_t^u , X_t^u , H_t^u , and J_t^u .

Figure 5 shows how the computer code mentioned above is cast into the bivariate SUTSE with Italian Industrial Production Index for Capital Goods and the German Truck Toll Index.

The same system may be applied to Italian Industrial Production Index of Intermediate Goods and the Truck Toll Index. As stochastic regressors, we use the commercial vehicle registration in

Germany and the commercial vehicle registration in Italy. In this last case, both regressors are available at time $t + 1$, while in the case for capital goods only the level of production of cars in Germany is present. The same gearings are applied for the system about non-durable goods and natural gas for industrial use. We use as stochastic regressors consumption of electricity and commercial vehicle registration in Italy always available at time $t + 1$. The Kalman filter log likelihood estimation does not force us to estimate the whole concatenated system. Once the bivariate SUTSE is estimated, we only have to combine this system with the star system and finally apply the Kalman Filter's smoothing recursion (see Gómez [14]) to obtain the new forecasts with stochastic exogenous regressors for Italy and Germany.

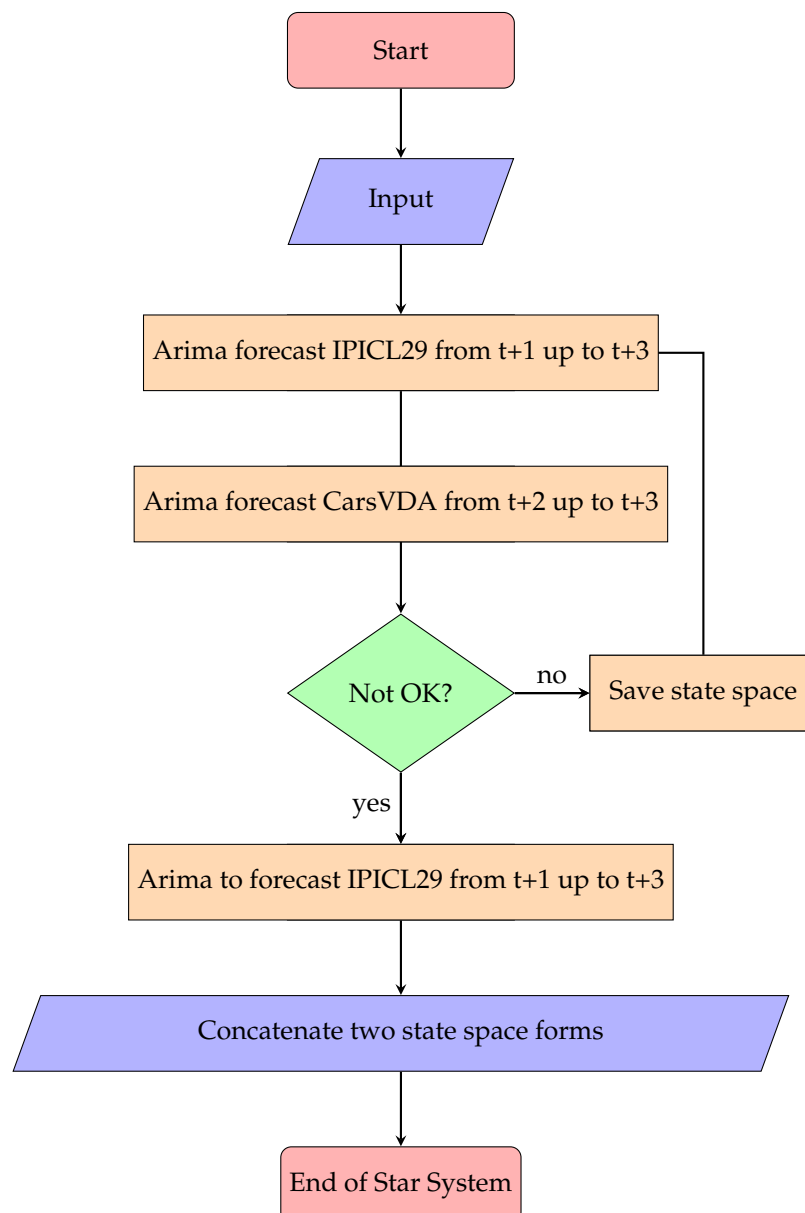


Figure 4. Star System.

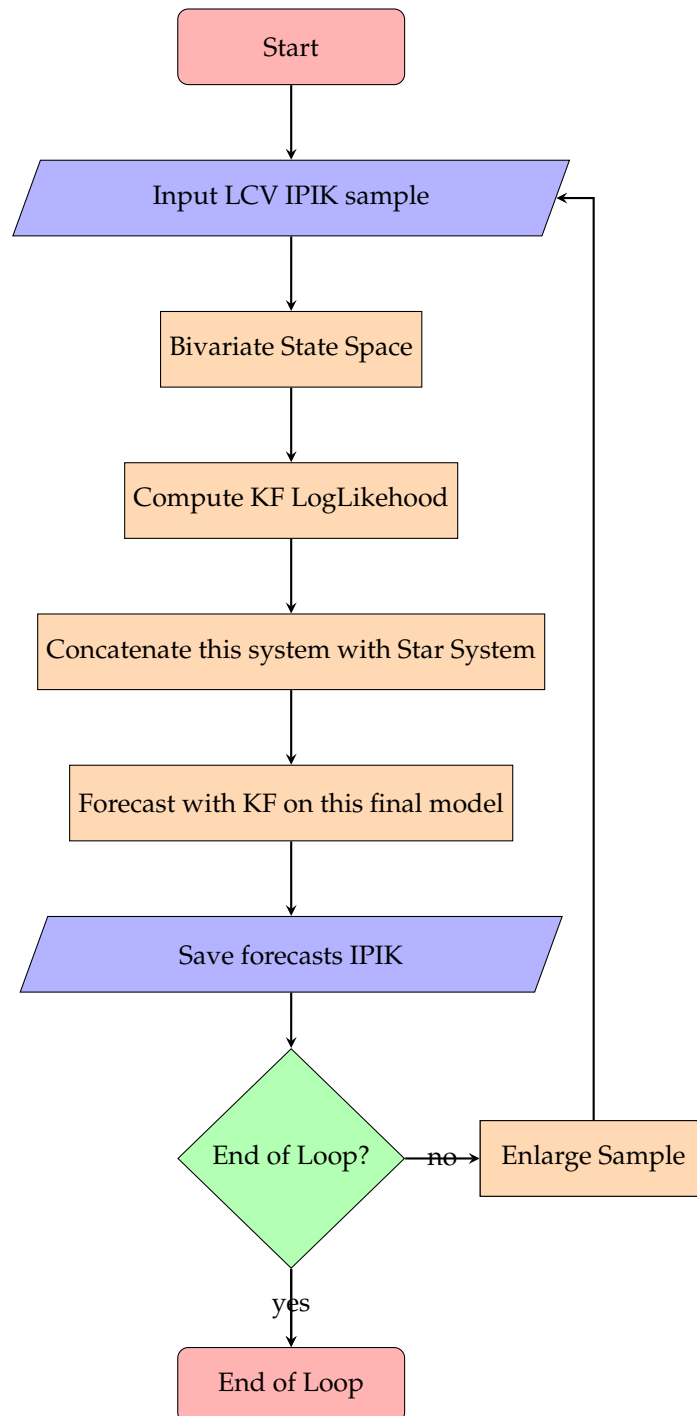


Figure 5. Concatenate Seemingly Unrelated Time Series Equations with star system.

3.3. Mixing Stochastic Regressors with Barebone Model due to Failure in Convergence

Sometimes it happens that the star system (see Figure 4) fails to converge due divergence between the two series. For instance, production of cars and trucks in Italy during the month of August is less labor-intensive than in Germany. Consequently, in certain months, such as August 2009, in a period of crisis, the divergence between German production of cars and the Italian twin is even more marked and the Kalman filter algorithm fails to converge in the star system. During the recursive estimation in the case of failure of convergence, a rollback strategy to the barebone model has been implemented. The same strategy applies to the enhanced barebone model for intermediate goods and for non-durable goods.

3.4. Long Weekends

During some months, there are fixed holidays on Tuesdays or Thursdays. A typical example is the Immaculate Conception holiday on the 8th of December. Many workers requested to go on holiday on the 7th of December 2015. In this way, they used four days (Saturday, the 5th of December; Sunday, the 6th of December; Monday, the 7th of December; and the day of the Immaculate Conception). The raw data about the Industrial Production Index are affected by such a slowdown on Monday, the 7th of December. Nevertheless, ISTAT does not consider the group of people who do not work during that day. The calendar and seasonal adjustment is strict. Alternatively, most of the times, the calendar adjusted time series are more predictable than gross ones.

Time series not affected by strikes or unexpected holidays could be predicted more easily than gross ones. To tackle this problem, on the calendar from the 1st of January of 2001 for the following days are traced:

- New Year's Day
- Epiphany: the 6th of January
- Italian Republic day: the 2nd of June
- Immaculate Conception: the 8th of December
- International Workers' Day: the 1st of May

To trace the long weekends, the following strategy is proposed:

- For New Year's Day, Epiphany, Italian Republic day, and Immaculate Conception, find the years when they happen on Tuesdays and Thursday.
- Create a binary dummy variable to inform the model there will be a discrepancy between what it expects and what happens in the real world and place it on Mondays or Fridays for daily data.
- For the 1st of May, let the computer decide if the month of April is affected when Labour Day happens on Tuesday or Thursday. The same algorithm is applied to New Year's Day.
- Compute dummy variables for the months over the years detected from the procedure mentioned above.

3.5. Long Weekends and Airline Model with No Logarithmic Transformation

To exploit the aforementioned dummies (see Section 3.4), I inserted them into a popular airline model with no logarithmic transformation.

$$(1 - B)(1 - B^{12})y_t = (1 - \theta B)(1 - \Theta B^{12})\epsilon_t + \beta X_t \quad (17)$$

where B is the backshift operator and X_t contains number of workings days in Italy on monthly basis and the whole set of long weekend dummies. Table A19 shows the final results at the end of the sample. This naive model is inserted into the forecasting competition (see Section 7).

3.6. Benchmark Model, MAE and RMSE over Seasonal Differences

$$\Delta_{12}\log(y_t) = \sum_{j=1}^3 \Delta_{12}\gamma_j \log(y_{t-j}) + \epsilon_t \quad (18)$$

Equation (18) describes the autoregressive benchmark used by Bulligan et al. [1]. The Mean Absolute Error (MAE) is computed over $\Delta_{12}\log(y_t)$. It is given by the following formula:

$$MAE = \frac{1}{n} \sum_{j=1}^n |\Delta_{12}\log(y_t) - \Delta_{12}\log(\hat{y}_t)| \quad (19)$$

where \hat{y} is the forecast by the model chosen and n is given by the end of the sample. To ease the comparison, the following ratio is used during the recursive experiment:

$$I_{MAE} = \frac{MAE_{DIS}}{MAE_{BEN}} \quad (20)$$

where MAE_{BEN} is approximately Equation (18) and MAE_{DIS} is given by the model chosen for the disaggregated component. Unless specified, MAE_{DIS} uses the barebone model shown in Section 3.1.

On the same path, the Root Mean Squared Error (RMSE) is again computed over $\Delta_{12} \log(y_t)$ and is given by the following formula:

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n [\Delta_{12} \log(y_t) - \Delta_{12} \log(\hat{y}_t)]^2} \quad (21)$$

The ratio between the disaggregated model forecast and the benchmark model is used:

$$I_{RMSE} = \frac{RMSE_{DIS}}{RMSE_{BEN}} \quad (22)$$

3.7. Tests of Predictive Accuracy

As Bruno et al. [4] for h-step ahead forecasts, we use the modified Diebold–Mariano statistic:

$$DM^* = \left(\frac{n+1-2h+n^{-1}h(h-1)}{n} \right)^{\frac{1}{2}} \frac{\bar{d}}{\sqrt{n^{-1}2\pi f_d(0)}} \quad (23)$$

where $n = 36$, $\bar{d} = n^{-1} \sum_{t=1}^n d_t$, $d_t = g(e_{1t}) - g(e_{2t})$ is an arbitrary function of the forecasting errors from model $i \in \{1, 2\}$, and $\widehat{f_d(0)}$ is a consistent estimate of the zero-frequency spectral density of d_t .

When performing tests of forecast encompassing, d_t becomes $d_t = e_{1t}(e_{1t} - e_{2t})$ (see [16]) under the alternative, Forecast 1 could be improved by incorporating some of the features present in Forecast 2. When comparing forecasting accuracy, in this paper, we use $d_t = |e_{1t}| - |e_{2t}|$ and $d_t = e_{1t}^2 - e_{2t}^2$.

Albuquerque [17] showed that: “The Hansen’s Test of Superior Predictive Ability can be addressed by testing the null hypothesis that the benchmark is not inferior to any alternative forecast. We seek a test of the null hypothesis that the benchmark is not inferior to any of the alternatives.” In other words,

$$d_{k,t} = (d_{1t} \dots d_{mt}) \quad (24)$$

where the vector $d_{k,t}$ is the vector of stacked relative alternatives.

Hansen test works under the assumption that model k is better than the benchmark if and only if equation $E(d_{k,t}) > 0$.

The Hansen [18] superior predictive ability test assumes that d_t is (strictly) stationary and second finite moment (to be possible to apply the Central Limit Theorem).

In this case, the asymptotic null distribution is:

$$n^{\frac{1}{2}} \hat{d} \sim N_{k,t}(0, \hat{\Omega})$$

At this point, all essential aspects of our framework are identical to those of [19] Reality Check (RC). White, H. [19] proceeded by constructing the RC from the test statistic:

$$T_n^{RC} = \max(n^{\frac{1}{2}} \hat{d}_1, \dots, \hat{d}_m)$$

where \hat{d}_k is the sample mean associated with the k th model and Ω is a consistent estimator of the variance and covariance matrix.

Berardi et al. ([20]) summarized the model confidence set procedure (see Hansen et al. [9]): The procedure starts from an initial set of models M^0 of dimension m encompassing all the model specifications $\hat{M}^0 \dots$ and delivers for a given confidence level $1 - \alpha$, a smaller set, the superior set of models, SSM, $\hat{M}_{1-\alpha}^*$ of dimension $m^* \leq m$. The best scenario is when the final set consists of a single model, i.e., $m^* = 1$. Formally, let $d_{ij,t}$ denote the loss differential between models i and j :

$$d_{ij,t} = l_{i,t} - l_{j,t}, i, j = 1 \dots m, t = 1, \dots, n \quad (25)$$

and let

$$d_{i,\cdot,t} = (m-1) \sum_{j \in M} d_{ij,t}, i = \dots, m \quad (26)$$

be the simple loss of model i relative to any other model j at time t . The EPA hypothesis for a given set of models M can be formulated in two alternative ways:

$$H_{0,M} : c_{ij} = 0, \text{ for all } i, j = 1, 2, \dots, m \quad H_{A,M} : c_{ij} \neq 0, \text{ for some } i, j = 1, 2, \dots, m \quad (27)$$

or

$$H_{0,M} : c_{ij} \neq 0, \text{ for all } i = 1, 2, \dots, m \quad H_{A,M} : c_{ij} = 0, \text{ for some } i = 1, 2, \dots, m \quad (28)$$

where $c_{ij} = E(d_{ij})$ and $c_i = E(d_i)$ are assumed to be finite and not time dependent. According to Hansen et al. ([9]), to test the two hypotheses above, the following two statistics are constructed:

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{var}(\bar{d}_{ij})}} \quad \text{and} \quad t_i = \frac{\bar{d}_i}{\sqrt{\widehat{var}(\bar{d}_i)}} \quad i, j \in M \quad (29)$$

where $\bar{d}_i = (m-1)^{-1} \sum_{j \in M} \bar{d}_{ij}$ is the simple loss of the i th model relative to the average losses across models in the set M , and $\bar{d}_{ij} = (m)^{-1} \sum_{t=1}^m d_{ij,t}$, t measures the relative sample loss between the i th and j th models, while $\widehat{var}(\bar{d}_{i,\cdot})$ and $\widehat{var}(\bar{d}_{ij})$ are bootstrapped estimates of $var(\bar{d}_{i,\cdot})$ and $var(\bar{d}_{ij})$, respectively. As discussed by Hansen et al. (see [9]), the two EPA null hypotheses presented in Equation (28) map naturally into the two test statistics:

$$T_{R,M} = \max_{i,j \in M} |t_{i,j}| \quad \text{and} \quad T_{max,M} = \max_{i \in M} t_i, \quad (30)$$

where $t_{i,j}$ and t_i are defined in Equation (29). The test statistics defined in Equation (30) can be used to test the two hypotheses in Equation (28), respectively. Since the asymptotic distribution of the two tests statistic is nonstandard, the relevant distribution under the null hypothesis is estimated using a bootstrap procedure similar to that used to estimate $var(\bar{d}_{i,\cdot})$ and $var(\bar{d}_{ij})$. The MCS procedure consists of a sequential testing procedure that eliminates the worst model at each step, until the hypothesis of equal predictive ability (EPA) is accepted for all the models belonging to the SSM. The choice to eliminate the worst model was carried using an elimination rule that is coherent with the statistic test defined in Equation (29):

$$e_{max,M} = \operatorname{argmax}_{i \in M} \frac{\bar{d}_{i,\cdot}}{\widehat{var}(\bar{d}_{i,\cdot})} \quad \text{and} \quad e_{R,M} = \operatorname{argmax}_{i \in M} \left\{ \sup_{j \in M} \frac{\bar{d}_{ij}}{\sqrt{\widehat{var}(\bar{d}_{ij})}} \right\} \quad (31)$$

respectively.

In summary, the MCS procedure to obtain the SSM consists of the following step:

- Step 1. Set $M = M_0$.
- Step 2. Test for EPA hypothesis: if EPA is accepted, terminate the algorithm and set $M_{1-\alpha}^* = M$, otherwise use the elimination rules defined in Equation (31) to determine the worst model.

Step 3. Remove the worst model, and go to Step 2.

This paper uses the MCS procedure (see Hansen et al. [9]) and test for superior predictive accuracy of Hansen (see [18]) implemented by Sheppard (see [21]) available for MATLAB.

4. VARMA Applied to the Whole Index

This section describes the model applied to the whole index in VARMA state–space form. The German Truck Toll Index is backcast using a naive reconstruction from 2001:01 to 2004:12. To this series is assigned the same seasonal growth rate of the consolidated published time series of Industrial Production Index for Intermediate Goods in Germany. In this way, endogenous variables can be employed in the following time series:

1. Industrial Production Index in Italy
2. Consumption of electricity in Italy
3. Truck Toll Index in Germany

Unfortunately, the time series data on natural gas for industrial use are not available before 2005:01. It is not possible to compute a naive backcast due to lack of information up to 2001:01. Since the last two variables have always been available at time $t + 1$, they are inserted in state–space form of the Kalman filter. The following general procedure is employed with an expanding window starting from 2001:01 to 2008:01. The last one is expanded at every step of the recursive estimation up to 2018:12. The mandatory steps to estimate this type of VARMA model are given below:

1. Identify the VARMA model with an iterative likelihood-ratio test procedure (see Gómez [12] for an analytical description).
2. Carry out a preliminary estimation of the parameters by the Hannan–Rissanen method (see Gómez [12] and Hannan et al. [22]).
3. Refine the estimation using the conditional method described by Lütkepohl (see [23]), Reinsel [24]) and Gómez (see [12,14] for the implementation).
4. Cast the new parameters into state–space form.
5. Use the Kalman filter for maximum likelihood estimation.
6. Perform diagnostic testing.
7. Estimate forecast using the information about other endogenous variables available at time $t+1$.

The aforementioned procedure can be applied to any VARMA process of any given dimension. This simplified VARMA model is used to compare the forecast with the aggregated model by means of the model confidence set (see Hansen et al. [9]). The Supplementary Materials presents two videos (see Supplementary Video S1 and Supplementary Video S2) that show the evolution of the p -values of the multivariate Q-statistic (see Gómez [12]) applied to vectorized residuals and squared vectorized residuals over 36 lags applied to the basic model. A similar model that does not exploit public available data (i.e., railway transportation data provided by Trenitalia Cargo and monthly level of temperatures) was described by Ventura et al. (see [25]) and does not exploit Kalman filtering. Their model is summarized by the following nested models:

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=0}^q \gamma_j' z_{t-j} + \epsilon_t \quad (32)$$

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=0}^q \gamma_j' z_{t-j} + \zeta_j' F_t + \eta_t \quad (33)$$

where F_t represents $R \times 1$ vectors of factors obtained via cross-validation, and ζ is a conformable coefficient vector. Since the block intermediate plus capital and electricity account roughly for seventy per cent in terms of weights about the whole index (see Figure 2), this study focuses only on the survey data concerning non-durable goods. In this way, it focuses on ninety per cent of all the information

available on the web for the whole index from January 2001 up to now. The factors are not extracted and all the raw data on surveys are inserted as they are published (not seasonally adjusted, as by Costantini [5] or log-linearised as by Bruno et al. [4] and, again, by Ventura et al. [25]). The Italian Industrial Production Index, the Truck Toll Index and the consumption of electricity are expressed as logarithms. Preliminary analyses show that the presence of unit roots at some seasonal frequencies cannot be excluded. Accordingly, all variables are transformed through seasonal differencing. Finally, the first difference filter is applied to seasonal differenced variables to achieve stationarity for energy and Truck Toll Index and Italian Industrial Production Index. Table 2 presents the survey data list as in the work of Costantini ([5]) concerning non-durable goods. These data are used in first difference according to the well-known general representation of the VARMA model:

$$\Phi(B)Y_t = \Theta(B)A_t + \beta X_t, \quad (34)$$

$$Y = \begin{bmatrix} Y1 \\ Y2 \end{bmatrix} \quad (35)$$

where $Y1 = \Delta\Delta_{12}\log(y_t)$ is the vector containing the variables listed in Table 2 for the hard data and $Y1 = \Delta y_t$ for the survey data about non-durable goods. Finally, X_t contains $\Delta\Delta_{12}\log(wd_t)$, where wd_t is the number of workings days in Italy and a couple of level shifts detected by SSSMATLAB for Italian Industrial Production Index. These thirteen endogenous variables are estimated recursively over the three-step horizon. Step 1 includes studying all the information already available for the other twelve variables. For the other two steps, the forecast is unconditioned. Figures 6–8 show the results of the Durbin–Watson test on autocorrelation, the Harvey test on heteroskedasticity and the Bowman–Shenton test for normality testing. It can be seen that the Durbin–Watson is largely above five per cent for the whole sample, and the heteroskedasticity test rejects the null hypothesis of no heteroskedasticity of the residuals at the five per cent level of significance for the whole of 2013 and the result of the Bowman–Shenton test on normality testing is negative, rejecting the null hypothesis at five per cent level of significance starting from the beginning of 2013 up to the end of the sample. Likewise, global crisis is hardly detected around August 2009 and, again, the test has roughly one year of rejection between the second half of 2011 and the first half of 2012.

Table 2. Information sources of data for VARMA using survey data.

Name	Series	Sources	Freq.	Time ^a
IPI	1	ISTAT	Monthly	-
GigaWatt Electricity Consumption	1	TERNA	Monthly	+35+40 days
Truck Toll Mileage Index in Germany	1	DESTATIS	Monthly	+35+40 days
Climate Index (non-durable goods)	1	ISTAT	Monthly	+35+40 days
General economy expectations (non-durable goods)	1	ISTAT	Monthly	+35+40 days
Assessment of orders (non-durable goods)	1	ISTAT	Monthly	+35+40 days
Assessment of domestic orders (non-durable goods)	1	ISTAT	Monthly	+35+40 days
Assessment of export orders (non-durable goods)	1	ISTAT	Monthly	+35+40 days
Production growth (non-durable goods)	1	ISTAT	Monthly	+35+40 days
Assessment of stocks of finished products (non-durable goods)	1	ISTAT	Monthly	+35+40 days
Orders expectations (non-durable goods)	1	ISTAT	Monthly	+35+40 days
Production expectations (non-durable goods)	1	ISTAT	Monthly	+35+40 days
Price expectations (non-durable goods)	1	ISTAT	Monthly	+35+40 days

^a Publication corresponds to amounts of days before the official release on ISTAT website.

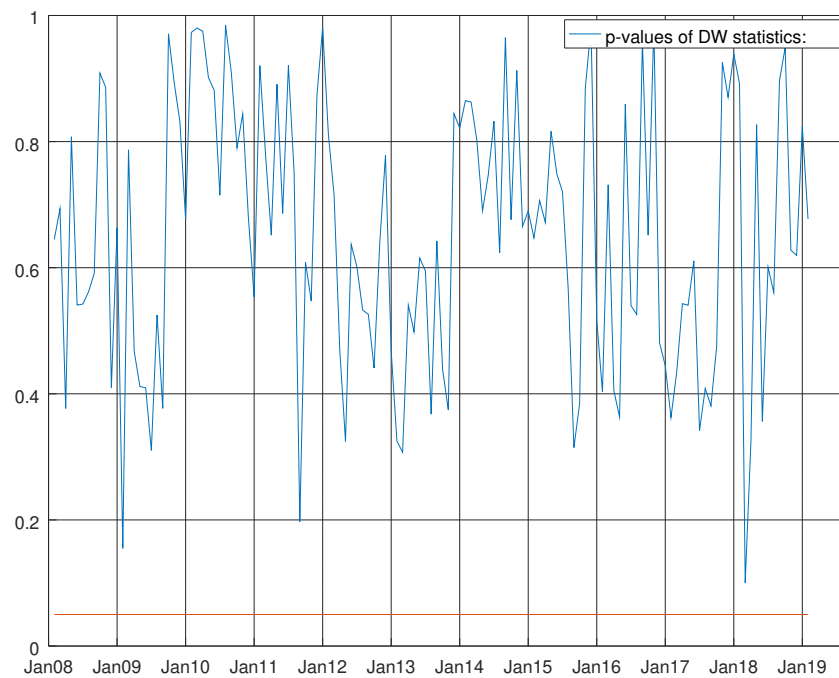


Figure 6. Graph showing the p -values of the Durbin–Watson test for the VARMA model using the hard data only. The null hypothesis states that residuals are not autocorrelated.

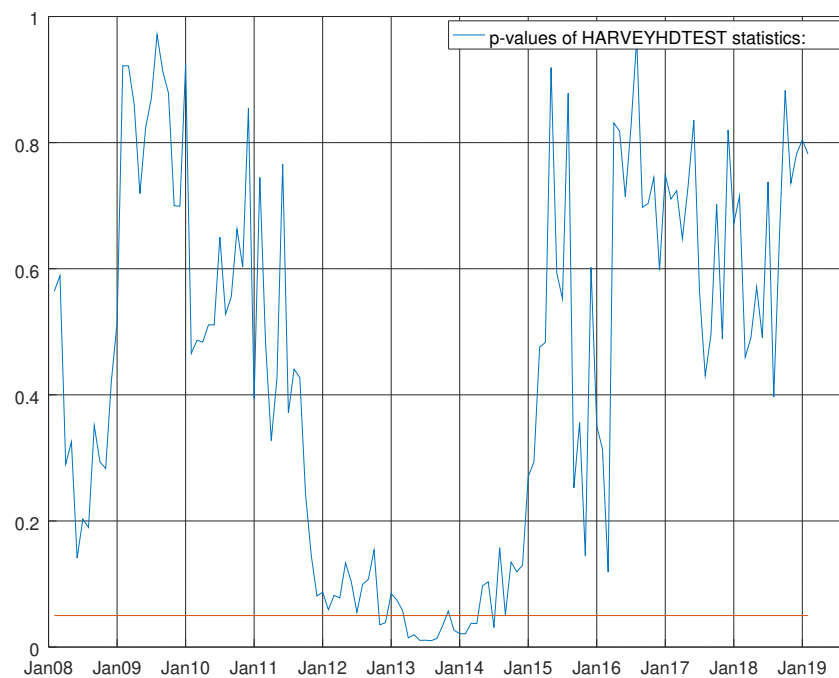


Figure 7. Graph showing the p -values of the Harvey heteroskedasticity test for the VARMA model using the hard data only. The null hypothesis states that the residuals are not heteroskedastic.

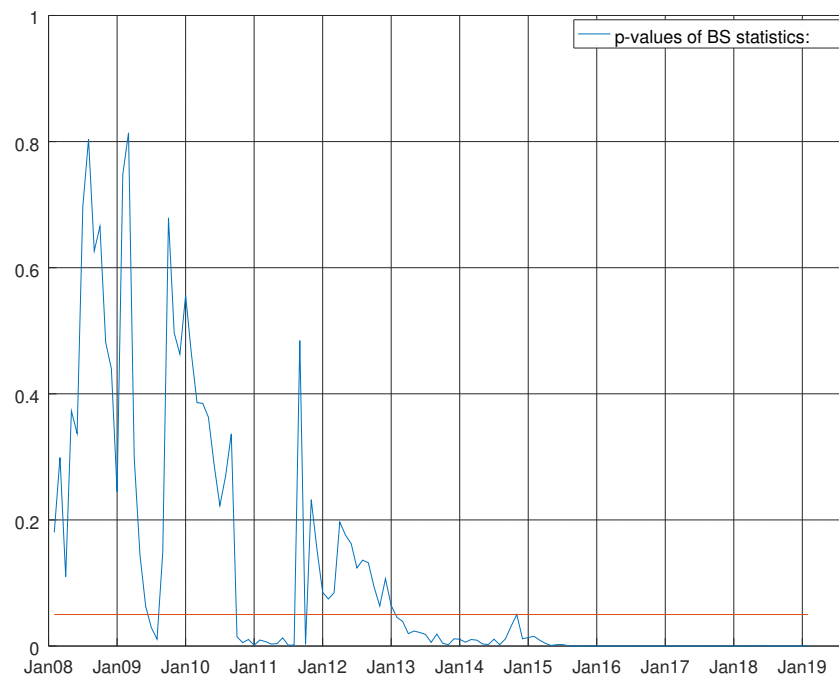


Figure 8. Graph showing the p -values of the Bowman–Shenton test for the VARMA model using the hard data only. The null hypothesis states that errors are normally distributed.

Adding the whole set of endogenous variables for non-durable goods (soft data) from the ISTAT survey apparently does not improve the overall picture (see Table 3). In fact, Table 4 shows that the modified Diebold–Mariano test (see Harvey et al. [16]) strongly rejects the null hypothesis of equal forecast accuracy compared to the benchmark of Bulligan et al. ([1]) on the first and third steps. In addition, at the second step, the p -value of the encompassing test strongly rejects the null hypothesis in August 2009, a peak of absolute error of the benchmark is observed. The diagnostic does not show an improvement with respect of the basic VARMA model using the hard data. Figure 9 shows the DW test for this VARMA using both hard and soft data for non-durable goods. Figure 10 shows the results of Normality testing. Figure 11 shows the Harvey test using both hard data and soft data for non-durable goods.

Table 3. Comparison of mean absolute error and root mean squared error with expanding window applied to the VARMA using only hard data and both hard data and soft data, respectively (i.e., ISTAT survey), about the non-durable goods. The forecast windows expands itself from 2008:01 to 2018:12. The ratio between MAEs and RMSEs with respect to the benchmark are shown. A value less than one indicates an improvement over the benchmark.

TYPE OF RATIO	Step 1		Step 2		Step 3	
	VARMA	VARMAPLUSND	VARMA	VARMAPLUSND	VARMA	VARMAPLUSND
IMAE	0.2749	0.6361	0.3142	0.9186	0.3089	0.8245
IMAE-2015	0.5316	1.0636	0.5356	1.5651	0.6034	1.4227
IRMSE	0.2370	0.6228	0.2765	0.9792	0.2670	0.7521
IRMSE-2015	0.4268	1.0627	0.5110	1.6598	0.5447	1.4713

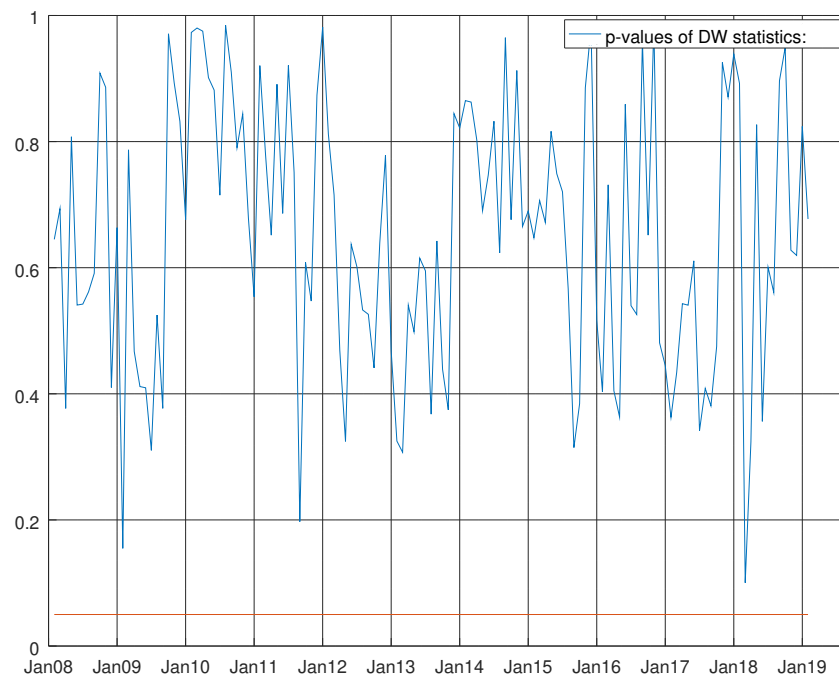


Figure 9. Graph showing the p -values of the Durbin–Watson test for the VARMA model using both hard and soft data for non-durable goods. The null hypothesis states that residuals are not autocorrelated.

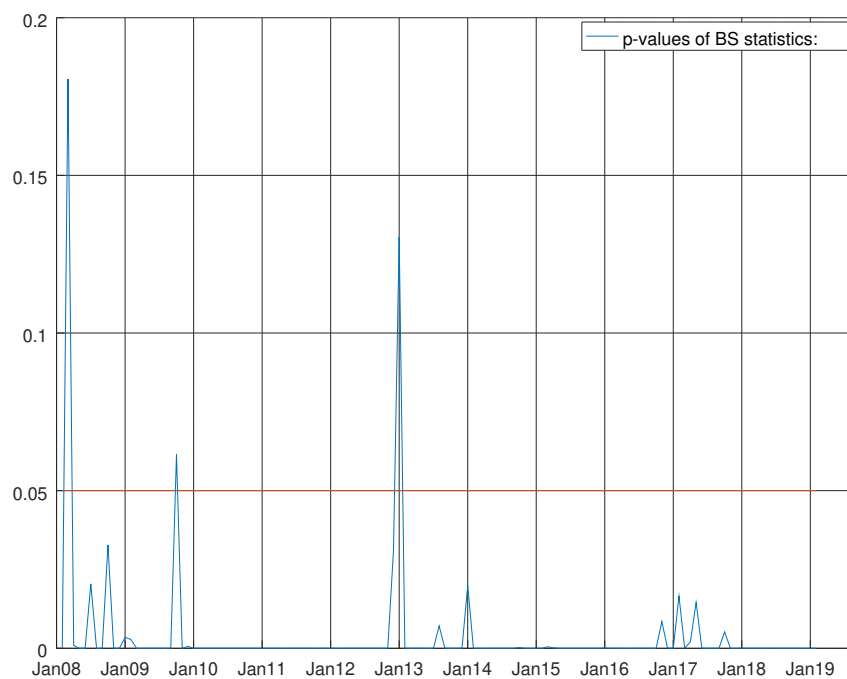


Figure 10. Graph showing the p -values of the Bowman–Shenton test for the VARMA model using both hard and soft data for non-durable goods. The null hypothesis states that errors are normally distributed.

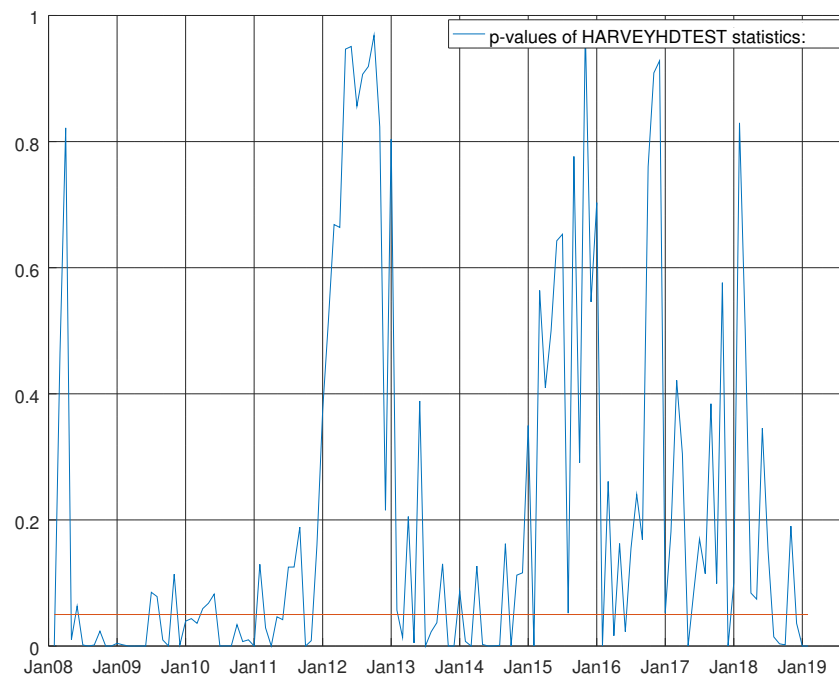


Figure 11. Graph showing the p -values of the Harvey heteroskedasticity test for the VARMA model using both hard and soft data for non-durable goods. The null hypothesis states that the residuals are not heteroskedastic.

Table 4. Tests of equal predictive accuracy and test of forecast encompassing of VARMA using only hard data and both hard data and soft data, respectively (i.e., ISTAT survey), for non-durable goods. The forecast window expands itself from 2008:01 to 2018:12. The three tests are considered at the end of the sample and at the maximum value of forecast benchmark, i.e., August 2009. The p -values near to one show that the null hypothesis of equal predictive accuracy have been accepted with respect to the benchmark of Bulligan et al. For encompass test, the null hypothesis is that the forecast of the benchmark encompasses the forecast of the VARMA.

MMDTEST-SUBSAMPLE-LOSS-FUNCTION	Step 1		Step 2		Step 3	
	VARMA	VARMAPLUSND	VARMA	VARMAPLUSND	VARMA	VARMAPLUSND
MMD-0801-1812-ABSERR	0.9250	0.0037	0.7183	0.1788	0.5040	0.0621
MMD-0801-1812-ERR2	0.5850	0.0000	0.4887	0.2381	0.7913	0.0205
MMD-0801-0908-ABSERR	0.7730	0.0149	0.3122	0.2703	0.6828	0.0795
MMD-0801-0908-ERR2	0.6770	0.0301	0.1723	0.7801	0.5070	0.0368
MMD-0801-0908-ENC	0.4300	0.0017	0.4224	0.0258	0.0258	0.0082
MMD-0801-1812-ENC	0.0000	0.0000	0.8011	0.3929	0.0021	0.0000

Table 5 shows how the second VARMA greatly benefits from the information already available at time $t + 1$ for hard and soft data (i.e., all the p -values are below the five per cent significance). At time $t + 2$, this gap is evident to strongly accept the null hypothesis that VARMA using just the hard data encompasses the model that uses both hard and soft data for non-durable goods. At time $t + 3$, the gap between the huge VARMA and hard data VARMA shrinks again. Again, at $t + 1$, the null hypothesis of encompassing both the peak of the crisis (i.e., August 2009 under two per cent significance) and the end of the sample (below one per cent significance) are rejected.

Table 5. Tests of equal predictive accuracy and test of forecast encompassing of VARMA using only hard data and both hard data and soft data, respectively, (i.e., ISTAT survey) for non-durable goods. The forecast windows expands itself from 2008:01 to 2018:12. The three tests are considered at the end of the sample and at the maximum value of forecast benchmark, i.e., August 2009. The p -values near one show that null hypothesis of equal predictive accuracy between the VARMA that uses only hard data and the VARMA that uses hard data plus soft data for non-durable goods has been accepted. For encompass test, the null hypothesis is that the forecast of the VARMA using only hard data encompasses the forecast of the VARMA using only hard data and soft data for non-durable goods.

MMDTEST-SUBSAMPLE-LOSS-FUNCTION	Step 1	Step 2	Step 3
	VARMA vs. VARMAPLUSND	VARMA vs. VARMAPLUSND	VARMA vs. VARMAPLUSND
MMD-0801-1812-ABSERR	0.0013	0.27661	0.1283
MMD-0801-1812-ERR2	0.0046	0.2889	0.0281
MMD-0801-0908-ABSERR	0.0238	0.1981	0.0800
MMD-0801-0908-ERR2	0.0349	0.4956	0.0548
MMD-0801-0908-ENC	0.0018	0.0173	0.0120
MMD-0801-1812-ENC	0.0000	0.8061	0.0000

5. Recursive Diagnostics of Disaggregated Models

This section presents the results of diagnostics related to the distinct disaggregated model used. For intermediate goods, we consider the regressors for following time series: (1) working days in Italy and Germany; (2) a linear trend; (3) a series of outliers and the level shift identified on the overall sample before the computer started the recursion by SSMMATLAB plus a level shift to inform the computer about our naive backcast for the period 2001:01–2004:12; (4) the level of registration of commercial vehicles in Italy; and (5) the level of of registration of commercial vehicles in Germany.

Figure 12 plots the difference between the recursive Bayesian information criterion (see Gómez [12] for a description of the diagnostics and Schwarz [26]) of the barebone model for intermediate goods and the enhanced model for intermediate goods. The Bayesian information criterion of the enhanced model is always inferior.

At the beginning of the recursive experiment, between January 2008 and January 2009, the enhanced model did not show its own supremacy. From the beginning of 2009, the situation constantly improved in favour of the enhanced model. In other words, starting from 2009, the new regressors had an upward contribution waving around a linear trend. This hints the use of the enhanced model as a long-run solution due to the growing importance of commercial vehicles registration in both countries to understand the long-run evolution of industrial production. On the basis of this evidence, we now focus on the diagnostics of this enhanced model for intermediate goods. Table A8 shows the final results at the end of the sample (December 2018). We see that the regressor coefficients of the stochastic regressors are statistically significant, and we strongly reject the null hypothesis on the basis of the t -values over time. The Supplementary Materials present a video (Supplementary Video S3) showing how the aforementioned p -values of the stochastic coefficients are constant and stable. Surprisingly, the impact of working days in Germany is superior to the impact of working days in Italy on the Italian Industrial Production Index of Intermediate Goods. For the Truck Toll Index, the difference is less outstanding. The coefficient of the linear trend shows a negative impact over the Italian data while its twin for the Truck Index shows a positive impact. This again does not come as a surprise given the persistent upward trend of the German data. Similar conclusions might be drawn from the coefficients of commercial vehicles registrations for both the countries. The Bowman–Shenton test (see Bowman et al. [27] and Figure 13), with unique exceptions towards the end of 2008 and the end of 2017, is always above the five per cent rejection area of the null hypothesis.

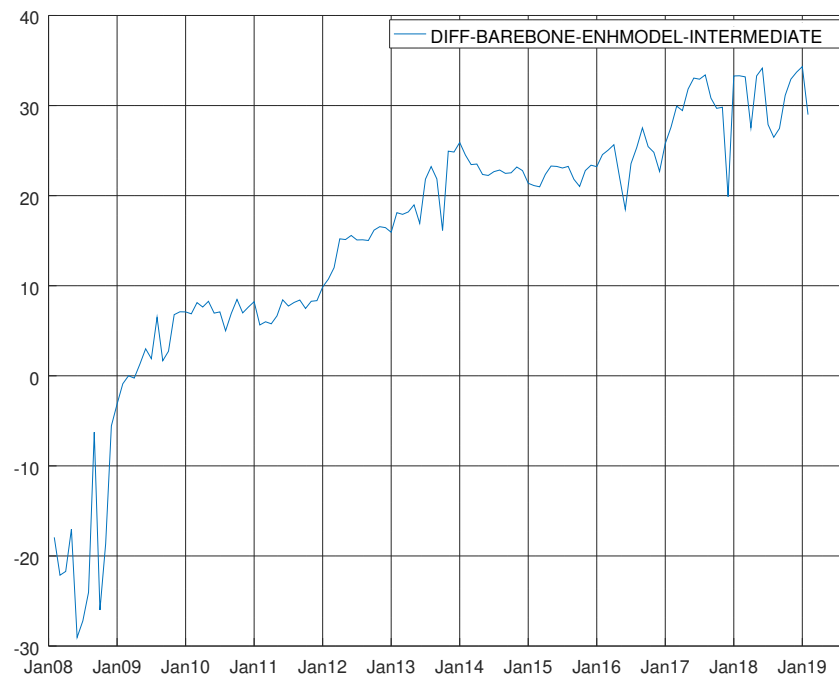


Figure 12. Graph showing the difference between the Bayesian information criterion of the barebone model and the enhanced model for intermediate goods. A negative value means that BB model prevails.

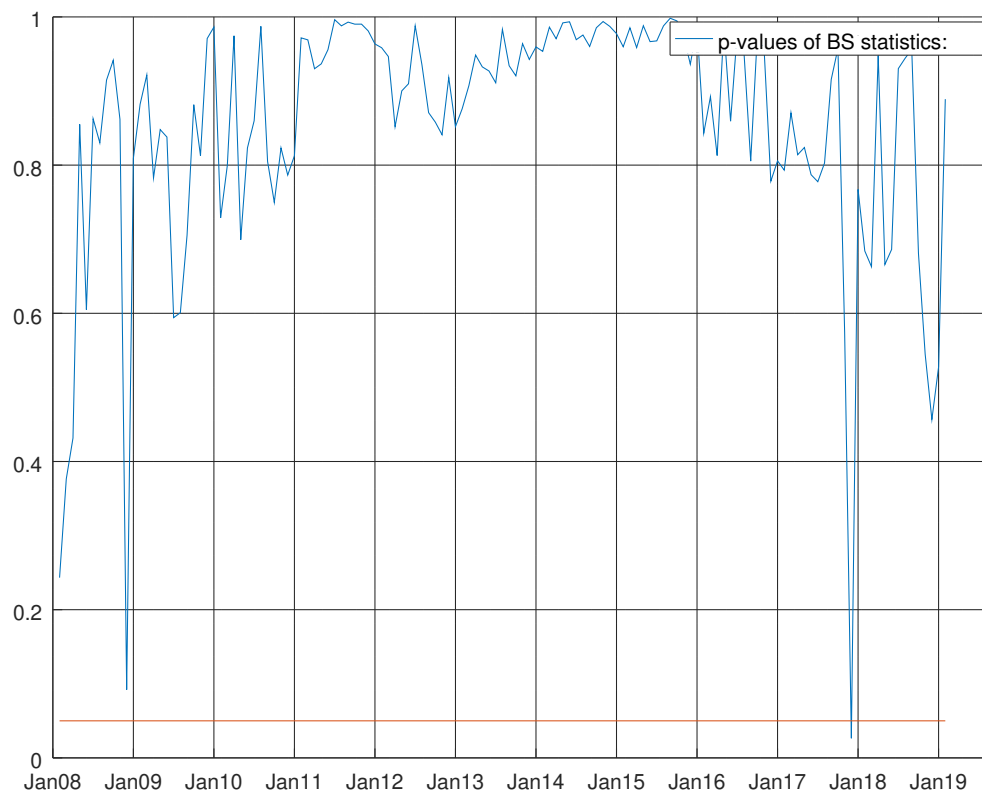


Figure 13. Graph showing the p -values of the Bowman-Shenton test for the enhanced model for intermediate goods. The null hypothesis states that errors are normally distributed.

Figure 14 shows the p -values of the Durbin–Watson statistics (see Durbin et al. [28,29]) for the enhanced barebone model of intermediate goods. At the bottom of the graph, the straight red line denotes 0.05 line significance. The null hypothesis of no autocorrelation of the residual was strongly accepted during the whole experiment. The results of heteroskedasticity test of Harvey are more unstable at the beginning of the experiment (see Figure 15) (see Harvey [13], p. 259).

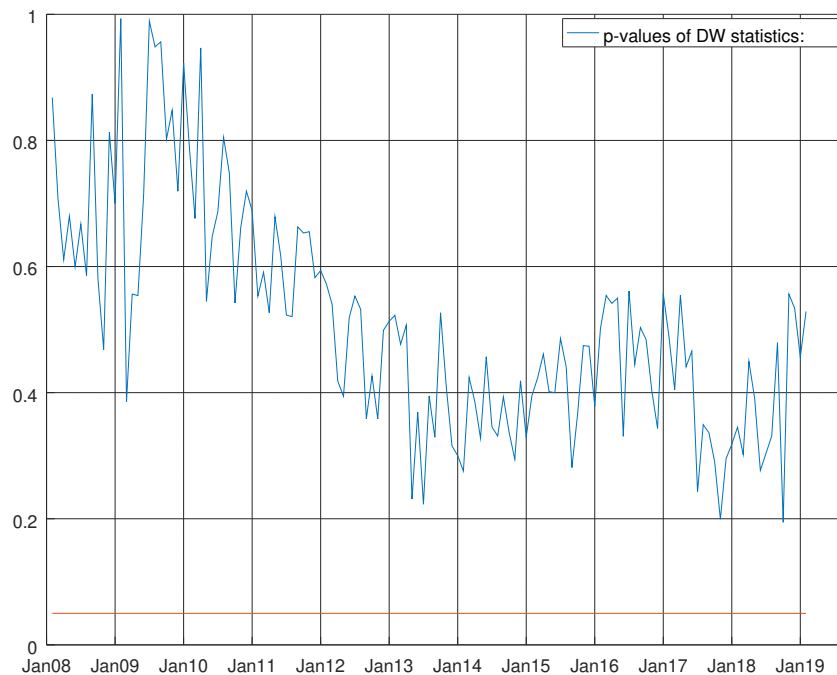


Figure 14. Graph showing the p -values of the Durbin–Watson test for the enhanced model for intermediate goods. The null hypothesis states that residuals are not autocorrelated.

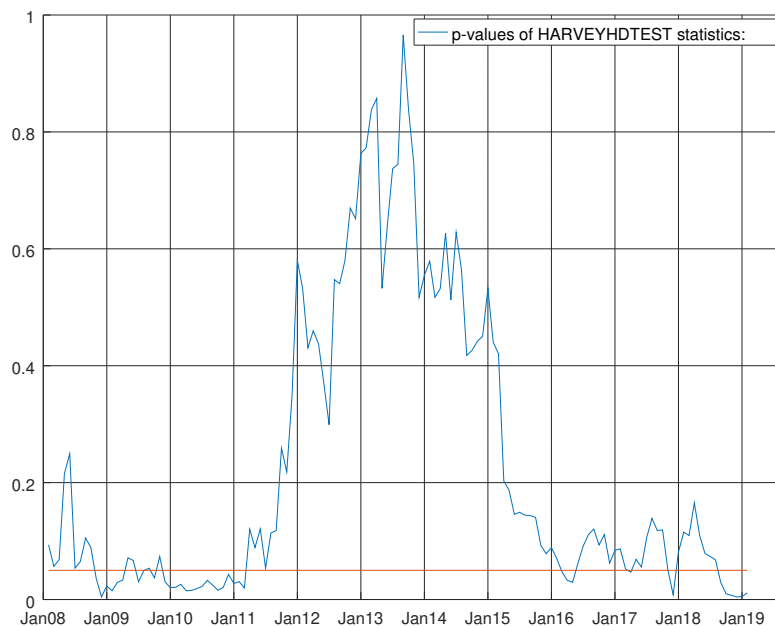


Figure 15. Graph showing the p -values of the Harvey heteroskedasticity test for the enhanced model for intermediate goods. The null hypothesis states that the residuals are not heteroskedastic.

Due to global economic crisis, the test is rejected at five per cent significance level for the years between 2009 and 2011. Nevertheless, the model began recovering from 2012 with an upward trend and then waved roughly from 2016 below the 0.2 line. Using the sample autocorrelations of the residuals and squared residuals, it is possible to show the evolution of Q-statistics for over 36 lags. The Supplementary Materials present two videos (Supplementary Video S4 and Supplementary Video S5) to show the movements of the p -values for both the tests over the 0.05 significance. It is possible to ensure that both the tests are hardly rejected over 36 lags and over time after the global crisis.

Again, the enhanced model shows its own supremacy in terms of the Bayesian information criterion for capital goods.

A direct comparison of Figures 12 and 16 shows how the difference in terms of the Bayesian information criterion is even more marked for capital goods than for intermediate goods. Even in this case, the enhanced model is the focus of this study. Harvey's test for heteroskedasticity has hardly been rejected from January 2008 up to the second half of 2016 (see Figure 17). The p -value of this test still waves around the significance red line for the remainder of the sample. Even more satisfying are the results of Bowman–Shenton statistics (see Bowman et al. [27] and Gómez [12]) over the whole experiment (see Figure 18). On the same line, the lack of autocorrelation between the residuals on the basis of the Durbin–Watson's p -value is depicted in Figure 19. In the Supplementary Materials, the flow of the Q-statistics over the three dimensions applied to the residual and the squared residuals can be seen (i.e., Supplementary Video S6, Supplementary Video S7). Even in this case, we can see that, in most instances, the p -values waved far from the bottom. Table A10 shows the values of the parameters estimated by the Kalman filter by maximum likelihood at the end of the sample. Both endogenous variables are more significantly affected by the level shift detected from 2009:02 than the other ones previously detected by SSSMATLAB. The coefficient about working days can almost be swapped in terms of value between the level for Italy and Germany. The working days in Germany have a positive effect over the Italian data while the working days in Italy have a slight negative effect over the German data. This can be explained by geographical reasons. Many firms concerning capital goods located in the north of Italy may be connected with German firms in the south of Germany. Thus, when Italians work more because they have more working days in a month, they may influence German workers. In addition, the linear trend has a stronger effect in Germany than in Italy. The contribution of commercial vehicle registration in Italy is more marked for German data than for Italian data. The picture is reversed for German data. The Supplementary Materials provide the evolution of p -values and t -statistics over time to check the parameters' stability (i.e., Supplementary Video S8). It can be seen that the overall picture of the parameters is stable over time.

Table A14 displays the results of maximum likelihood estimation at the end of the sample for the barebone model for durable goods. The level shifts from 2008:1 are in the group of the most significant coefficients among the selected variables. Figures 20–22 summarize the results of the recursive Durbin–Watson test for autocorrelation, Bowman–Shenton test for normality and Harvey test for heteroskedasticity. Once again, the null hypothesis is not rejected. In term of weights (see Figure 2), the results concerning the Italian Industrial Production Index of Non-Durable Goods are more relevant. Figure 23 shows the difference between the Bayesian information criterion of the barebone model and the enhanced model for non-durable goods. The difference between the other aforementioned figures (see Figures 12 and 16) about the intermediate and capital is more evident. Nevertheless, even for non-durable goods, the enhanced model shows a better BIC for the major part of the experiment. Two of the three usual tests for normality (see Figures 24–26) do not show a unique convincing picture for the last three years. Table A15 shows the results at the end of the sample for the enhanced model for non-durable goods. It can be seen that the impact and the statistical significance of consumption of electricity is higher for Industrial Production Index of Non-durable Goods than for natural gas for industrial use. Conversely, the impact of commercial vehicle registration in Italy is higher in natural gas for industrial use. The video in the Supplementary Materials concerning the

p -values and t -statistics' evolution over time (see Supplementary Video S9) shows more movements than the enhanced model for intermediate goods and than the model for capital goods.

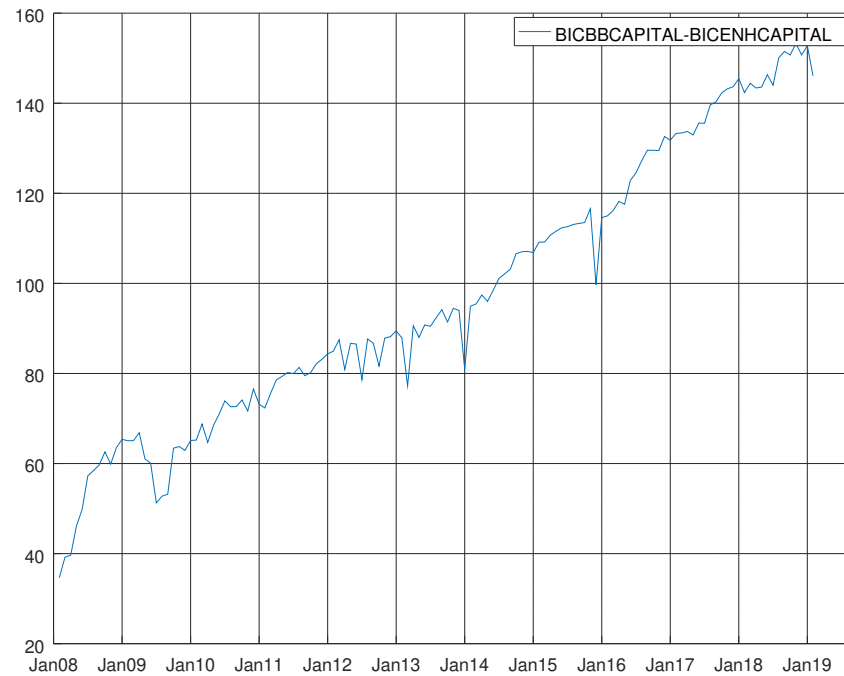


Figure 16. Graph showing the difference between the Bayesian information criterion of the barebone model and the enhanced model for capital goods.

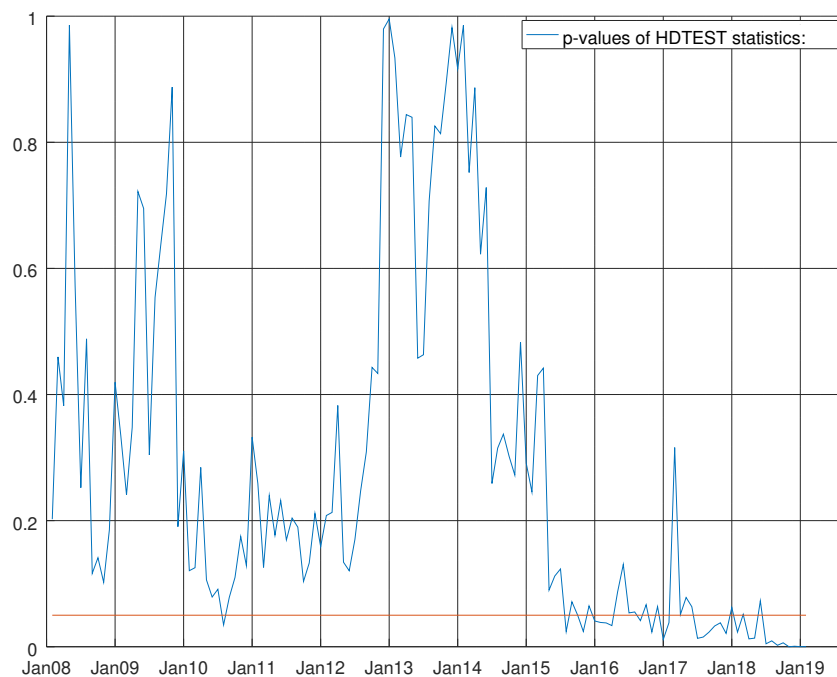


Figure 17. Graph showing the p -values of the Harvey heteroskedasticity test for the enhanced model for capital goods. The null hypothesis states that the residuals are not heteroskedastic.

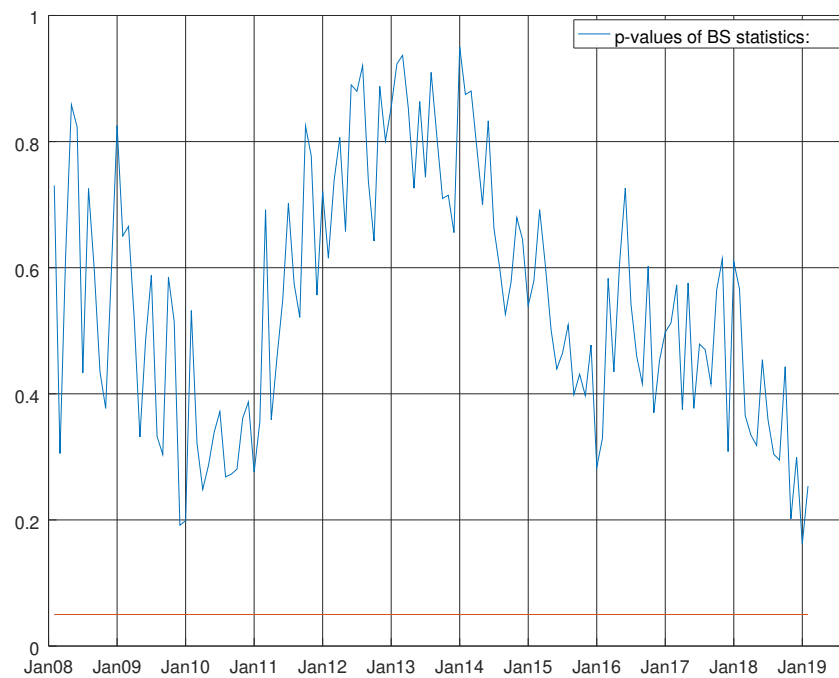


Figure 18. Graph showing the p -values of the Bowman–Shenton test for the enhanced model for capital goods. The null hypothesis states that errors are normally distributed.

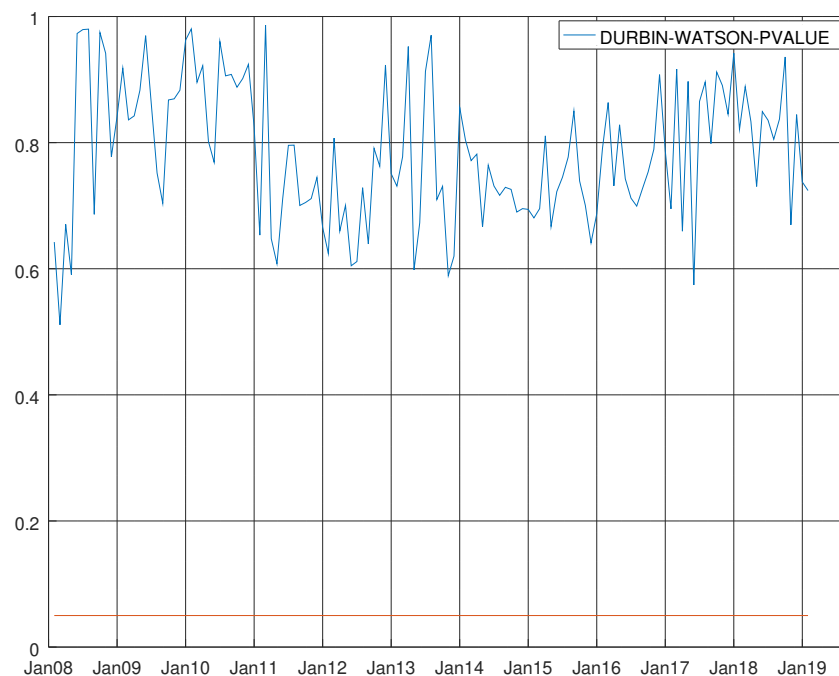


Figure 19. Graph showing the p -values of the Durbin–Watson test for the enhanced model for capital goods. The null hypothesis states that residuals are not autocorrelated.

Figure 25 shows the p -values of Durbin–Watson statistics (see Durbin et al. [28,29]) for the enhanced barebone model of non-durable goods. At the bottom of the graph, the straight red line

denotes 0.05 line significance. The null hypothesis of no autocorrelation of the residual is strongly accepted during the whole experiment. Towards the end of the second half of 2016 (see Figure 26), the results of heteroskedasticity test of Harvey (see Harvey [13], p. 259) are less stable than the past.

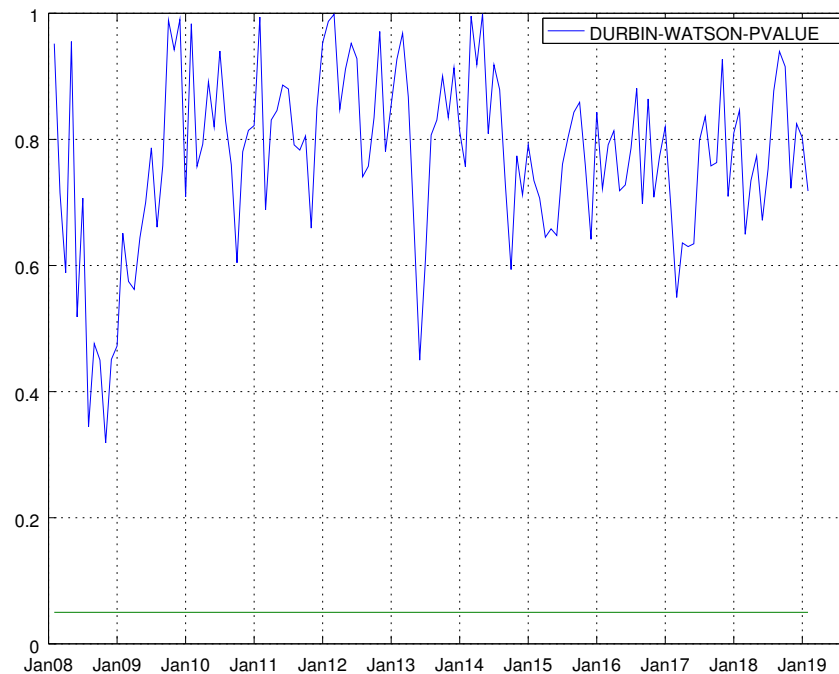


Figure 20. Graph showing the p -values of the Durbin–Watson test for the barebone model for durable goods. The null hypothesis states that residuals are not autocorrelated.

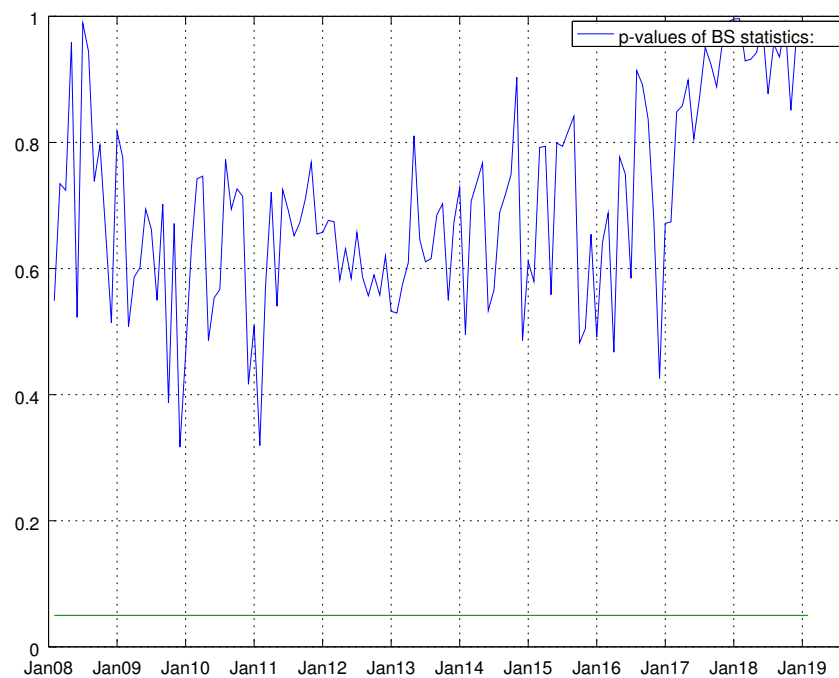


Figure 21. Graph showing the p -values of the Bowman–Shenton test for the barebone model for durable goods. The null hypothesis states that errors are normally distributed.

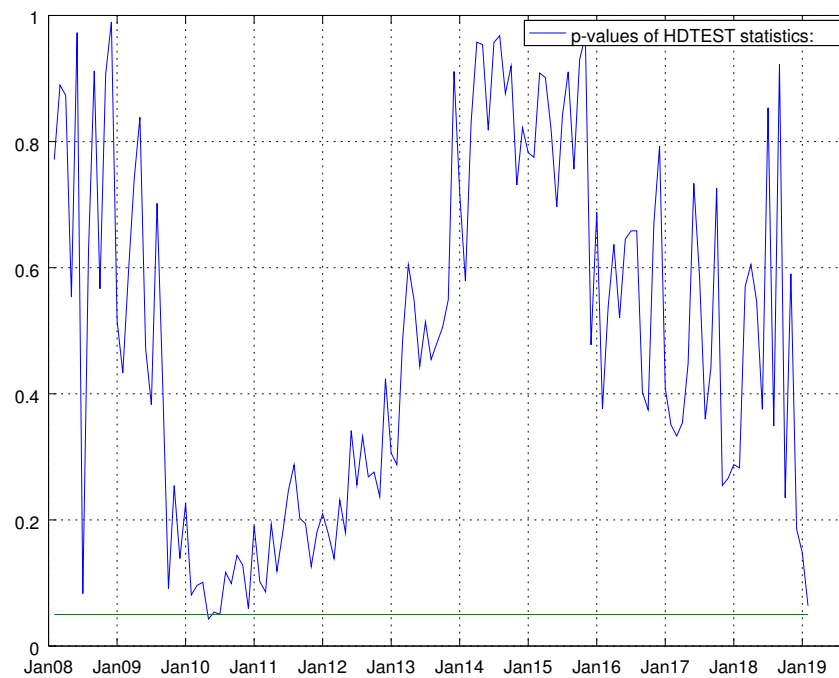


Figure 22. Graph shows the p -values of the Harvey heteroskedasticity test for the barebone model for durable goods. The null hypothesis states that the residuals are not heteroskedastic.

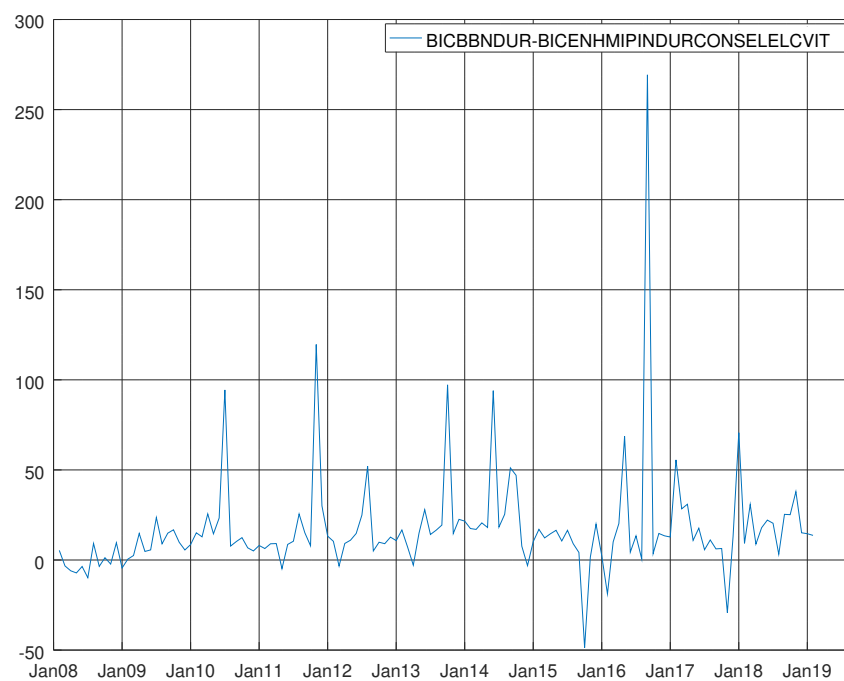


Figure 23. Graph showing the difference between the Bayesian information criterion of the barebone model and the enhanced model for non-durable goods.

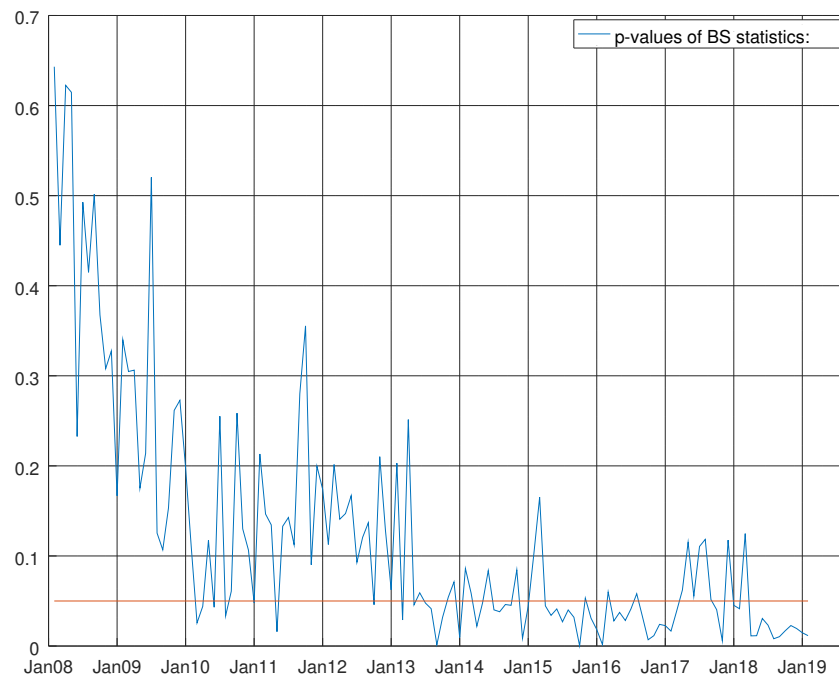


Figure 24. Graph showing the p -values of the Bowman–Shenton test for the enhanced model for non-durable goods. The null hypothesis states that errors are normally distributed.

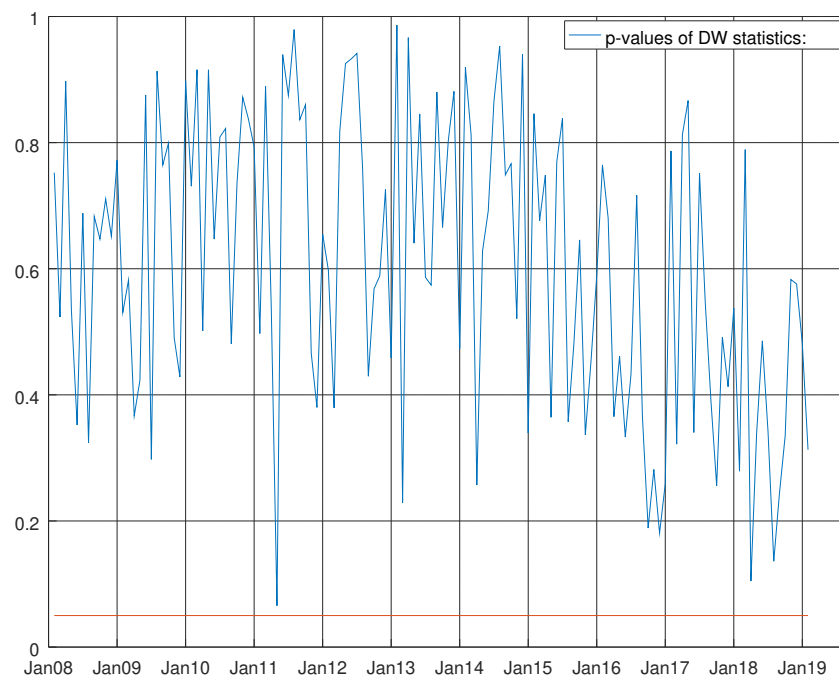


Figure 25. Graph shows the p -values of the Durbin–Watson test for the enhanced model for non-durable goods. The null hypothesis states that residuals are not autocorrelated.

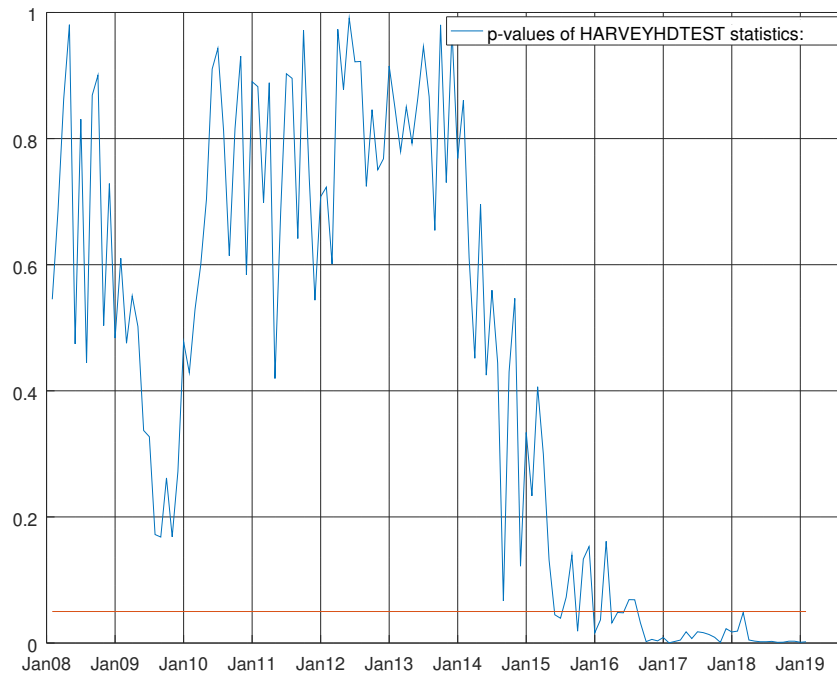


Figure 26. Graph showing the p -values of the Harvey heteroskedasticity test for the enhanced model for non-durable goods. The null hypothesis states that the residuals are not heteroskedastic.

6. Forecasting Study and Evaluation

6.1. Kalman Filter and State-Space Models

As indicated in Chapter 4 of Gómez [14], using the measurement equation and the transition equation shown in Section 3.1, the Kalman filter is given by the following recursion

$$E_t = Y_t - H_t \hat{x}_{t|t-1}, \Sigma_t = H_t P_t H_t' + R_t K_t = (F_t P_t H_t' + G_t S_t) \Sigma_t^{-1}, \hat{x}_{t+1|t} = F_t \hat{x}_{t|t-1} + K_t E_t \quad (36)$$

$$P_{t+1} = F_t P_t F_t' + G_t Q_t G_t' - (F_t P_t H_t' + G_t S_t)' = (F_t + K_t H_t) P_t F_t' + (G_t Q_t - K_t S_t') G_t', \quad (37)$$

initialized with $\hat{x}_{1|0} = a$ and $P_1 = \Omega$. Again, Gómez [14] showed the log-likelihood

$$l(Y) = \text{constant} - \frac{1}{2} \left\{ \frac{1}{\sigma^2} \sum_{t=1}^n E_t' \Sigma_t^{-1} E_t + \sum_{t=1}^n \ln |\sigma^2 \Sigma_t| \right\} \quad (38)$$

and that maximizing the log-likelihood above is equivalent to minimizing the following non-linear sum of squares

$$S = \left(\prod_{t=1}^n |\sigma_t| \frac{1}{2np} \right) \left(\sum_{t=1}^n e_t' e_t \right) \left(\prod_{t=1}^n |\Sigma_t| \frac{1}{2np} \right) \quad (39)$$

The specialized software described by Gómez [12,15] handles this special form of S where $e_t = \Sigma_t^{-\frac{1}{2}} E_t$.

6.2. Smoothing

Gómez [14] (Section 4.10) showed the following recursions are used for $t = n, \dots, 1$

$$(L_t, \lambda_t) = Z_t' \Sigma_t^{-1} (E_t, e_t) + K_{p,t}' (L_{t+1}, \lambda_{t+1}), \Lambda_t = Z_t' \Sigma_t^{-1} Z_t + K_{p,t}' \Lambda_{t+1} K_{p,t} \quad (40)$$

where $K_{p,t} = T_t - K_t Z_t$, initialized with $(L_{n+1}, \lambda_{n+1}) = (0, 0)$ and $\Lambda_{n+1} = 0$. Letting $\hat{\alpha}_{t|n} = E(\alpha_t|Y)$ be the orthogonal projection of α_t onto Y and $\hat{P}_{t|n} = \text{MSE}(\hat{\alpha}_{t|n})$ for $t = n, \dots, 1$, the fixed interval smoother is given by

$$\hat{\alpha}_{t|n} = \left[\left(-U_t, \alpha_{t|t-1} \right) + P_t (L_t, \lambda_t) \right] (-\hat{\gamma}_{n+1}, 1)'$$

$$P_{t|n} = \left[P_{t|n} + (P_t L_t - U_t) \prod_{n+1} (P_t L_t - U_t)' \right] \hat{\sigma}^2$$

where $\hat{\gamma}_{n+1}$ and \prod_{n+1} are the GLS estimator of γ for Y and its MSE.

6.3. Some Reflections over the Period before December 2014

The period from January 2001 to December 2014 requires an ad-hoc treatment to compare the performance of the disaggregated composed model and a naive benchmark model, i.e., autoregressive model of order three over the log seasonal differences of the whole index.

For this, two issues need to be tackled:

- backcast consumption of electricity daily data from 2006 to 2013 exploiting the compressed natural gas for thermoelectricity daily data; and
- since the weight of every sub-component is not fixed for this period, e.g. from January 2015 up to present time (i.e., they were rebuilt in terms of growth rate at the cost of fixed weight base year structure), it is mandatory to compute the discrepancy between the data published and the data obtained, holding the structure of base year weight (i.e., 2015) constant.

$$\begin{aligned} IPI &= \alpha IPI_x + \beta IPI_y + \text{discrepancy} : \\ \text{discrepancy} &= \text{discrepancy}_\alpha + \text{discrepancy}_\beta; \\ \text{discrepancy}_\alpha &= \text{discrepancy} \cdot \alpha; \\ \text{discrepancy}_\beta &= \text{discrepancy} \cdot \beta \end{aligned} \quad (41)$$

6.4. Forecasting Competition for the Enhanced Model Applied to Intermediate Goods, Capital Goods and Non-Durable Goods

It can be argued that Equation (33) faces the following problems:

- It uses log transformation.
- It applies the difference operator twice (first time with respect to the season and the second with respect to the year).

See Section 3.1 and Gómez [12] for a description of SUTSE models (SSSMATLAB computer programs are available from the author upon special request).

Table 6 outlines the results of the model confidence set procedure (see Hansen et al. [9]) at the end of the sample using absolute error as the loss function. The supremacy of the enhanced model is shown in the last line of the table (see ENHTTMI). The difference with its competitors shrinks at Step 3. This is probably due to lack of information. This picture is confirmed looking at Table 7. Table 8 outlines the results of the model confidence set procedure at the end of the sample using squared error as the loss function. In this case, the difference with respect to the simple barebone model competitors (in particular, BBTMI and BBLCVSP) is stable over time. Table 9 shows the same results only at Steps 2 and 3. Table 10 shows a restricted group of model with similar results for capital goods. Once again, the choice of stochastic regressors enhanced the model in an upper position. This results is more important compared to the one shown in Table 11. On this table, the victory of the enhanced model is not evident on the sample. On the other hand we are interested if the victory of the enhanced model is statistically significant, so the results of the previous table are more convincing for us. Within

the ratios of root mean square errors (RMSE) with respect to the benchmark of Bulligan et al. ([1]), Table 12 shows that within the sample the difference is not so significant. Table 13 shows that the use of stochastic regressors over the three-step horizon is the winning choice on the side of model confidence set procedure for capital goods.

The picture is somewhat more linear for the enhanced model for non-durable goods using absolute error and quadratic error as loss function for model confidence set procedure (see Tables 14 and 15) and the IMAE (Table 16) and IRMSE (Table 17) tables.

Table 6. Model confidence set hierarchy loss function is the absolute error at the end of the sample for intermediate goods model comparison $\alpha = 0.05$, number of bootstrap replications = 5000 and block length = 12. The null hypothesis is that the average performance of the model in the row is as small as the minimum average performance across the remaining models. The alternative is that the minimum average loss across the remaining models is smaller than the average performance of the model in the row.

Index	Step 1	Step 2	Step 3
IPI-BEN-INTERM	0.0216	0.0102	0.0088
BBLCVDE	0.0362	0.3634	0.7236
BBLCVFR	0.0362	0.3634	0.8332
BBLCVSP	0.2184	0.3634	0.8332
BBTTMI	0.2184	0.5904	0.8332
ENHTTMI	1.0000	1.0000	1.0000

Table 7. Comparison of the ratios of mean absolute error for models for intermediate goods with an expanding window from 2008:01 to 2018:12.

Index	Step 1	Step 2	Step 3
BBLCVDE	0.4749	0.5284	0.5672
BBLCVFR	0.4535	0.5181	0.5537
BBLCVSP	0.4431	0.5180	0.5478
BBTTMI	0.4304	0.5134	0.5560
ENHTTMI	0.4345	0.4885	0.5362

Table 8. Model confidence set hierarchy loss function is the squared error at the end of the sample for intermediate goods model comparison $\alpha = 0.05$, number of bootstrap replications = 5000 and block length = 12. The null hypothesis is that the average performance of the model in the row is as small as the minimum average performance across the remaining models. The alternative is that the minimum average loss across the remaining models is smaller than the average performance of the model in the row.

Index	Step 1	Step 2	Step 3
IPI-BEN-INTERM	0.1490	0.1530	0.1414
BBLCVDE	0.1490	0.1530	0.1414
BBLCVFR	0.1490	0.2798	0.3096
BBLCVSP	0.3568	0.2798	0.3096
BBTTMI	0.3568	0.2798	0.3096
ENHTTMI	1.0000	1.0000	1.0000

Table 9. Comparison of the ratios of RMSE for models for intermediate goods with an expanding window from 2008:01 to 2018:12.

Index	Step 1	Step 2	Step 3
BBLCVDE	0.6603	0.7419	0.7993
BBLCVFR	0.6143	0.7166	0.7496
BBLCVSP	0.5835	0.6698	0.7009
BBTTMI	0.6323	0.7344	0.7774
ENHTTMI	0.6296	0.6539	0.7011

Table 10. Model confidence set hierarchy loss function is the absolute error at the end of the sample for capital goods model comparison $\alpha = 0.05$, number of bootstrap replications = 5000 and block length = 12. The null hypothesis is that the average performance of the model in the row is as small as the minimum average performance across the remaining models. The alternative is that the minimum average loss across the remaining models is smaller than the average performance of the model in the row.

Index	Step 1	Step 2	Step 3
IPI-BEN-CAPITAL	0.0040	0.0036	0.0010
BBLCVDE	0.6618	0.0356	0.3012
BBLCVFR	0.8222	0.9166	0.9820
BBLCVSP	0.9952	0.9328	0.9820
BBTMI	0.9952	0.9328	0.9820
ENHTTMI	1.0000	1.0000	1.0000

Table 11. Comparison of the ratios of absolute error for models for capital goods with an expanding window from 2008:01 to 2018:12.

Index	Step 1	Step 2	Step 3
BBLCVDE	0.4508	0.4943	0.5391
BBLCVFR	0.4543	0.4846	0.5327
BBLCVSP	0.4363	0.4924	0.5386
BBTTMI	0.4331	0.4907	0.5355
ENHTTMI	0.4344	0.5381	0.5755

Table 12. Comparison of the ratios of RMSE for models for capital goods with an expanding window from 2008:01 to 2018:12.

Index	Step 1	Step 2	Step 3
BBLCVDE	0.5287	0.5795	0.6065
BBLCVFR	0.5393	0.5690	0.5911
BBLCVSP	0.4606	0.5705	0.5558
BBTTMI	0.5350	0.5935	0.6137
ENHTTMI	0.5292	0.6261	0.6586

Table 13. Model Confidence Set Hierarchy Loss function is the squared error at the end of the sample for capital goods model comparison $\alpha = 0.05$, number of bootstrap replications = 5000 and block length = 12. The null hypothesis is that the average performance of the model in the row is as small as the minimum average performance across the remaining models. The alternative is that the minimum average loss across the remaining models is smaller than the average performance of the model in the row.

Index	Step 1	Step 2	Step 3
IPI-BEN-CAPITAL	0.1412	0.0770	0.0708
BBLCVDE	0.7124	0.0770	0.0708
BBLCVFR	0.7124	0.7062	0.6688
BBLCVSP	0.7124	0.7736	0.6688
BBTMI	0.7124	0.8882	0.6688
ENHTTMI	1.0000	1.0000	1.0000

Table 14. Model confidence set hierarchy loss function is the absolute error at the end of the sample for non-durable goods model comparison $\alpha = 0.05$, number of bootstrap replications = 5000 and block length = 12. The null hypothesis is that the average performance of the model in the row is as small as the minimum average performance across the remaining models. The alternative is that the minimum average loss across the remaining models is smaller than the performance of the model in the row.

Index	Step 1	Step 2	Step 3
IPI-BEN-NDUR	0.0000	0.0000	0.0000
BBSNAMINDUSTRIAL	0.0000	0.0000	0.0002
ENHSNAMINDUSTRIAL	1.0000	1.0000	1.0000

Table 15. Model confidence set hierarchy loss function is the squared error at the end of the sample for non-durable goods model comparison $\alpha = 0.05$, number of bootstrap replications = 5000 and block length = 12. The null hypothesis is that the average performance of the model in the row is as small as the minimum average performance across the remaining models. The alternative is that the minimum average loss across the remaining models is smaller than the average performance of the model in the row.

Index	Step 1	Step 2	Step 3
IPI-BEN-NDUR	0.0000	0.0000	0.0000
BBSNAMINDUSTRIAL	0.0000	0.0000	0.0002
ENHSNAMINDUSTRIAL	1.0000	1.0000	1.0000

Table 16. Comparison of the ratios of absolute error for models for capital goods with an expanding window from 2008:01 to 2018:12.

Index	Step 1	Step 2	Step 3
IPI-BEN-NDUR	1.0000	1.0000	1.0000
BBSNAMINDUSTRIAL	0.6513	0.7434	0.7251
ENHSNAMINDUSTRIAL	0.6407	0.5787	0.5964

Table 17. Comparison of the ratios of RMSE for models for capital goods with an expanding window from 2008:01 to 2018:12.

Index	Step 1	Step 2	Step 3
IPI-BEN-NDUR	1.0000	1.0000	1.0000
BBSNAMINDUSTRIAL	0.6292	0.7334	0.7211
ENHSNAMINDUSTRIAL	0.6286	0.6969	0.7053

7. Results and Discussion

Table 18 summarizes the performances of the models applied to the disaggregated components mentioned in Figure 2 using the IMAE mentioned in Section 3.7. Table 19 summarizes the results of the modified Diebold-Mariano test mentioned in Section 3.7 using as loss function absolute forecast error. On the same path Table 20 focuses on the the IMAE concerning the subsample from January 2015 up to December 2018, while Table 21 points out the results the modified Diebold-Mariano test mentioned in Section 3.7 using as loss function absolute forecast error for the subsample from January 2015 up to December 2018. Table 22 shows the performances of the models applied to the disaggregated components mentioned in Figure 2 using the IRMSE mentioned in Section 3.7. Table 23 focuses on the results the modified Diebold-Mariano test mentioned in Section 3.7 using as loss function quadratic forecast error. Again on the same path Table 24 focuses on the the IRMSE concerning the subsample from January 2015 up to December 2018, while Table 25 points out the results the modified Diebold-Mariano test mentioned in Section 3.7 using as loss function quadratic forecast error for the subsample from January 2015 up to December 2018.

An expanding window starting from January 2001 and moving from January 2008 to December 2018 is considered. (The dataset is updated on a monthly basis and is not prone to high revisions (see dati.istat.it/). It was downloaded on 8 February 2019.)

The disaggregated model outperforms the benchmark at Step 1 and is still competitive at Step 2 due to the information available from the energy sector and, at Step 3, still holds due to good performance of the intermediate, capital and non-durable goods. Disaggregating simply using the past does not help. The line IPI-BENAGG shows that aggregating the benchmarks applied to the sub-components does not change the situation. This concept is confirmed looking at the results shown in Tables 26–32. Using 25 days of data, we achieved better results at Step 2 and still competitive results due to an-hoc identification production on daily data (see Gómez et al. [11]). Tables 27, 29, 31 and 33 summarize the results of the modified Diebold–Mariano test mentioned in Section 3.7 for electricity.

We see that especially for the subsample from January 2015 to December 2018 the sub-component in charge of production of electricity can accept the null of equal forecast accuracy, especially at Step 2. Finally, it recovers at Step 3, where we may reject the null hypothesis of equal forecast accuracy at ten per cent significance. The low performance of IPI-PETROL is due to the lack of information for this sub-component. We can accept the null hypothesis of equal forecast accuracy with respect to the benchmark of using different loss functions only for durable goods. Given their low importance in terms of weights, the overall impact is negligible.

Table 18. Comparison of the ratios of mean absolute error with an expanding window from 2008:01 to 2018:12.

Index	Step 1	Step 2	Step 3
IPI-INT	0.4345	0.4885	0.5362
IPI-CAP	0.4200	0.5265	0.5692
IPI-DUR	0.4976	0.5006	0.5088
IPI-NDUR	0.6407	0.7196	0.7329
IPI-ENERGY	0.1498	0.5145	0.5328
IPI-AGG	0.3503	0.4254	0.4611
IPI-BENAGG	1.0019	1.0044	1.0035

Table 19. Modified Diebold–Mariano statistics for equality of forecast accuracy of two forecasts under general assumptions with an expanding window from 2008:01 to 2018:12. The null hypothesis is that the two methods have the same forecast accuracy. The loss function is the absolute error.

Index	Step 1	Step 2	Step 3
IPI-INT	0.0000	0.0000	0.0000
IPI-CAP	0.0000	0.0000	0.0000
IPI-DUR	0.0452	0.0539	0.0516
IPI-NDUR	0.0000	0.0000	0.0000
IPI-ENERGY	0.0000	0.0000	0.0000
IPI-AGG	0.0000	0.0000	0.0000

Table 20. Comparison of the ratio between mean absolute errors with an expanding window from 2015:01 to 2018:12.

Index	Step 1	Step 2	Step 3
IPI-INT	0.4373	0.5935	0.6499
IPI-CAP	0.3890	0.4412	0.4288
IPI-DUR	0.8689	0.9072	0.8852
IPI-NDUR	0.5974	0.6673	0.7049
IPI-ENERGY	0.1603	0.6020	0.6094
IPI-AGG	0.3590	0.5206	0.4957
IPI-BENAGG	1.0051	1.0180	1.0170

Table 21. Modified Diebold–Mariano statistics for the equality of forecast accuracy of two forecasts under general assumptions with an expanding window from 2015:01 to 2018:12. The null hypothesis is that the two methods have the same forecast accuracy. The Loss function is the absolute error.

Index	Step 1	Step 2	Step 3
IPI-INT	0.0001	0.0057	0.0179
IPI-CAP	0.0000	0.0000	0.0000
IPI-DUR	0.3469	0.4986	0.4446
IPI-NDUR	0.0000	0.0000	0.0000
IPI-ENERGY	0.0000	0.0081	0.0012
IPI-AGG	0.0000	0.0001	0.0002

Table 22. Comparison between RMSE ratios with an expanding window from 2008:01 to 2018:12.

Index	Step 1	Step 2	Step 3
IPI-INT	0.6296	0.6539	0.7011
IPI-CAP	0.5631	0.6695	0.7140
IPI-DUR	0.9298	0.9136	0.8500
IPI-NDUR	0.6286	0.6969	0.7053
IPI-ENERGY	0.1556	0.5361	0.5550
IPI-AGG	0.3917	0.4724	0.5249
IPI-BENAGG	0.9982	1.0036	1.0037

Table 23. Modified Diebold–Mariano statistics for the equality of forecast accuracy of two forecasts under general assumptions with an expanding window from 2008:01 to 2018:12. The null hypothesis is that the two methods have the same forecast accuracy. The loss function is the squared error.

Index	Step 1	Step 2	Step 3
IPI-INT	0.0389	0.0291	0.0338
IPI-CAP	0.0041	0.0268	0.0252
IPI-DUR	0.7985	0.8010	0.9780
IPI-NDUR	0.0000	0.0000	0.0000
IPI-ENERGY	0.0000	0.0000	0.0000
IPI-AGG	0.0000	0.0000	0.0000

Table 24. Comparison of mean ratio between RMSEs with an expanding window from 2015:01 to 2018:12.

Index	Step 1	Step 2	Step 3
IPI-INT	0.4517	0.6324	0.6783
IPI-CAP	0.4066	0.4357	0.4306
IPI-DUR	1.4071	1.4748	1.4073
IPI-NDUR	0.5891	0.6392	0.6439
IPI-ENERGY	0.1558	0.5921	0.6161
IPI-AGG	0.3642	0.5176	0.4926
IPI-BENAGG	0.9977	1.0048	1.0006

Table 25. Modified Diebold–Mariano statistics for the equality of forecast accuracy of two forecasts under general assumptions with an expanding window from 2015:01 to 2018:12. The null hypothesis is that the two methods have the same forecast accuracy. The loss function is the squared error.

Index	Step 1	Step 2	Step 3
IPI-INT	0.0000	0.0003	0.0005
IPI-CAP	0.0003	0.0146	0.0312
IPI-DUR	0.0001	0.0001	0.0001
IPI-NDUR	0.8694	0.7945	0.9430
IPI-ENERGY	0.0000	0.0000	0.0000
IPI-AGG	0.0002	0.0121	0.0053

Table 26. Comparison of ratios between mean absolute errors with an expanding window for electricity from 2008:01 to 2018:12.

Index	Step 1	Step 2	Step 3
IPI-EXTRA	0.3910	0.4073	0.7706
IPI-PETROL	0.6150	0.6712	0.7422
IPI-PRODELE	0.0075	0.5686	0.6056
IPI-CNG	0.1543	0.2073	0.7581
IPI-AGG	0.1498	0.5145	0.5328
IPI-BENAGG	0.9939	0.9830	0.9836

Table 27. Modified Diebold–Mariano statistics for the equality of forecast accuracy of two forecasts under general assumptions with expanding window from 2008:01 to 2018:12. The null hypothesis is that the two methods have the same forecast accuracy. The loss function is the absolute error.

Index	Step 1	Step 2	Step 3
IPI-EXTRA	0.0000	0.0000	0.0094
IPI-PETROL	0.0000	0.0000	0.0000
IPI-PRODELE	0.0000	0.0005	0.0000
IPI-CNG	0.0000	0.0000	0.0004
IPI-AGG	0.0000	0.0000	0.0000

Table 28. Comparison of the ratio between mean absolute errors with an expanding window for electricity from 2015:01 to 2018:12.

Index	Step 1	Step 2	Step 3
IPI-EXTRA	0.3707	0.3893	0.6791
IPI-PETROL	0.6763	0.7953	0.7441
IPI-PRODELE	0.0054	0.6713	0.7433
IPI-CNG	0.0961	0.1593	0.7501
IPI-AGG	0.1603	0.6020	0.6094
IPI-BENAGG	0.8733	0.8573	0.8575

Table 29. Modified Diebold–Mariano for the equality of forecast accuracy of two forecasts under general assumptions with an expanding window from 2015:01 to 2018:12. The null hypothesis is that the two methods have the same forecast accuracy. The loss function is the absolute error.

Index	Step 1	Step 2	Step 3
IPI-EXTRA	0.00000	0.00000	0.02050
IPI-PETROL	0.00950	0.06220	0.20580
IPI-PRODELE	0.00000	0.78530	0.03960
IPI-CNG	0.00000	0.00000	0.02340
IPI-AGG	0.00000	0.00810	0.00120

Table 30. Comparison of the ratio between RMSEs with an expanding window for electricity from 2008:01 to 2018:12.

Index	Step 1	Step 2	Step 3
IPI-EXTRA	0.3852	0.3967	0.8318
IPI-PETROL	0.6272	0.7061	0.7946
IPI-PRODELE	0.0100	0.4496	0.4778
IPI-CNG	0.0184	0.0235	0.0815
IPI-AGG	0.1556	0.5361	0.5550
IPI-BENAGG	1.0012	0.9986	0.9992

Table 31. Modified Diebold–Mariano statistics for the equality of forecast accuracy of two forecasts under general assumptions with an expanding window from 2008:01 to 2018:12. The null hypothesis is that the two methods have the same forecast accuracy. The loss function is the squared error.

Index	Step 1	Step 2	Step 3
IPI-EXTRA	0.0000	0.0000	0.1298
IPI-PETROL	0.0000	0.0000	0.0001
IPI-PRODELE	0.0000	0.0010	0.0000
IPI-CNG	0.0000	0.0000	0.0003
IPI-AGG	0.0000	0.0000	0.0000

Table 32. Comparison of the ratio between RMSEs with an expanding window for electricity from 2015:01 to 2018:12.

Index	Step 1	Step 2	Step 3
IPI-EXTRA	0.3330	0.3524	0.6130
IPI-PETROL	0.7224	0.7541	0.7532
IPI-PRODELE	0.0072	0.4659	0.5101
IPI-CNG	0.0105	0.0180	0.0787
IPI-AGG	0.1558	0.5921	0.6161
IPI-BENAGG	0.9129	0.9019	0.8997

Table 33. Modified Diebold–Mariano statistics for the equality of forecast accuracy of two forecasts under general assumptions with an expanding window from 2015:01 to 2018:12. The null hypothesis is that the two methods have the same forecast accuracy. The loss function is the squared error.

Index	Step 1	Step 2	Step 3
IPI-EXTRA	0.0022	0.0023	0.0164
IPI-PETROL	0.0212	0.1571	0.4489
IPI-PRODELE	0.0017	0.5629	0.0809
IPI-CNG	0.0000	0.0000	0.0163
IPI-AGG	0.0002	0.0121	0.0053

Finally, the model confidence set procedure described in Section 3.7 can be used to make comparison with all the models mentioned in this paper. Using the absolute forecast error as a loss function (see Table 34), the aggregated model on average is found to be the best, and there are no competitors. Using the quadratic forecast error as a loss function (see Table 35), the difference shrinks due to lack of information. Nevertheless, the aggregated model is still the best over the three-step horizon.

Table 34. Model confidence set hierarchy loss function is the absolute error at the end of the sample $\alpha = 0.05$, number of bootstrap replications = 5000 and block length = 12.

Index	Step 1	Step 2	Step 3
IPI-BEN	0.0010	0.0000	0.0000
IPI-AIRLINELW	0.0010	0.0000	0.0000
IPI-VARMABM	0.0010	0.0030	0.0010
IPI-VARMABMND	0.0010	0.0150	0.0150
IPI-AGG	1.0000	1.0000	1.0000

Table 35. Model confidence set hierarchy loss function is squared error at the end of the sample $\alpha = 0.05$, number of bootstrap replications = 5000 and block length = 12.

Index	Step 1	Step 2	Step 3
IPI-BEN	0.0566	0.0398	0.1012
IPI-AIRLINELW	0.0566	0.0692	0.1012
IPI-VARMABM	0.0566	0.0692	0.1012
IPI-VARMABMND	0.0566	0.0692	0.1012
IPI-AGG	1.0000	1.0000	1.0000

8. Conclusions

We evaluated the gains in the information from the highest disaggregation level at time $t + 1$ and time $t + 2$. Using statistical inference, we found that, for enhanced barebone models for intermediate goods, capital goods were the best in terms of superior predictive ability with respect to the set

benchmark as well as other combinations of the barebone model not using the German Truck Toll Index. The barebone model for durable goods is the weakest model compared to the set benchmark due to lack of information concerning this index. Nevertheless, its relative weight (four per cent) is negligible. We found that, using a quadratic forecast error loss function, we accept the null hypothesis of equal forecast accuracy between the enhanced model for non-durable goods and the benchmark in the subsample from January 2015 up to December 2018 (see Table 25). This result for non-durable goods is not confirmed (see Table 21). On average, the enhanced model for non-durable goods is still competitive over three steps horizon.

Of course, when the forecasting errors propagate over time, even a naive autoregressive model (see Section 3.5) is competitive with a disaggregated one (see Steps 2 and 3 of Table 35). Nevertheless, guessing the proper data generating process inside the main components allows us to shrink the forecast bands, even on the third-step horizon (see Tables 34 and 35), where some information is available.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2571-8800/2/4/33/s1>, Video S1: Evolution of Q-statistics and p-values over time and over 36 lags applied to vectorized residuals of Varma Basic Model , Video S2: Evolution of Q-statistics and p-values over time and over 36 lags applied to squared vectorized residuals of Varma Basic Model, Video S3: Evolution of t-statistics and p-values over time of regressors coefficients of Enhanced Model for intermediate goods, Video S4: Evolution of Q-statistics and p-values over time and over 36 lags applied to vectorized residuals of of Enhanced Model for intermediate goods , Video S5: Evolution of Q-statistics and p-values over time and over 36 lags applied to squared vectorized residuals of of Enhanced Model for intermediate goods, Video S6: Evolution of Q-statistics and p-values over time and over 36 lags applied to vectorized residuals of of Enhanced Model for capital goods , Video S7: Evolution of Q-statistics and p-values over time and over 36 lags applied to squared vectorized residuals of of Enhanced Model for capital goods, Video S8: Evolution of t-statistics and p-values over time of regressors coefficients of Enhanced Model for capital goods, Video S9: Evolution of t-statistics and p-values over time of regressors coefficients of Enhanced Model for non-durable goods.

Funding: This research received no external funding.

Acknowledgments: The views expressed in this paper are those of the author and do not represent those of the ISTAT. This article describes personal research in progress by the author and is published to elicit comments, to further debate and call for material collaboration in future developments of this research project. Thanks are given to all participants the 35th International Symposium on Forecasting–Riverside–California, U.S.A. Thanks are given to Fabio Bacchini and Giancarlo Bruno from Istat. I am grateful to Victor Gómez, statistician and technical advisor at the Spanish Ministry of Finance and Public Administrations in Madrid, for his constant effort to let people understand the inner workings of the Kalman filter.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IPI	Italian Industrial Production Index
LCV-SUM	Sum of the level of registration of light commercial van over Germany, Italy, France and Spain
LCV-GER	Level of registration of light commercial vehicles in Germany
LCV-ITA	Level of registration of light commercial vehicles in Italy
LCV-FRA	Level of registration of light commercial vehicles in France
LCV-SPAIN	Level of registration of light commercial vehicles in Spain
SUTSE	Seemingly unrelated time series equations
VARMA	Vector autoregressive moving average
IPI-AGG	Forecast obtained by aggregating sub index forecasts
IPI-AGGFROM201501	Forecast obtained by aggregating sub index forecasts from January 2015
IPI-AGG	Forecast obtained by aggregating naive benchmark sub index forecasts

IPI-INT	Italian Industrial Production Index of Intermediate goods
IPI-CAP	Italian Industrial Production Index of Capital Goods
IPI-DUR	Italian Industrial Production Index of Durable goods
IPI-NDUR	Italian Industrial Production Index of Non-durable goods
IPI-EXTRA	Italian Industrial Production Index of extraction of crude petroleum and natural gas
IPI-PETROL	Italian Industrial Production Index of manufacture of coke and refined petroleum products
IPI-PRODELE	Italian Industrial Production Index of electric power generation, transmission and distribution
IPI-CNG	Italian Industrial Production Index of manufacture of gas; distribution of gaseous fuels through mains
IPI-ENERGY	Italian Industrial Production Index of Electricity, gas, steam and air conditioning supply
PRODCARS	Level of cars produced in Germany
SNAM-IND	Level of M^3 of natural gas for industrial use
IPICL29	Italian Industrial Production Index of manufacture of motor vehicles, trailers and semi-trailers
CH4-PRODNAZ	Level of M^3 of natural gas produced in Italy
CH4-THERMO	Level of M^3 of natural gas for thermoelectric use
CH4-TRANSPORTED	Total level of M^3 of natural gas for transported on its own net
WDITA	Working days in Italy
WDGER	Working days in Germany
BIC	Bayesian information criterion
AIK	Akaike information criterion
TTMI	German Truck Toll Mileage Index
IPI-AIRLINELW	Airline model with no logarithmic transformation and workings days and dummy variables for long weekends
IPI-BEN	Benchmark model described in Bulligan et al. [1]
TC	Transitory Component
LS	Level Shift
AO	Anomalous Outlier

Appendix A

Appendix A.1

It is possible to download daily data from January 2013 to the present from www.terna.it and www.snam.it. On the one hand, we have daily data from 2006 to 2012 on the production of natural gas for thermoelectric use. On the other hand, only chunks of information are available for the consumption of electricity during this period in the past. We assume that in the remote past renewable sources of energy (e.g., wind, solar power, hydroelectricity, and geothermic source) had less relevance.

$$\Delta_7 \log(C_t) = \text{const}_t + \phi_1 \Delta_7 \log(C_{t-1}) + \phi_2 \Delta_7 \log(C_{t-2}) + \phi_3 \Delta_7 \log(C_{t-3}) + \beta_1 \Delta_7 \log(CH4_t) + \gamma D_t + \epsilon_t \quad (\text{A1})$$

Equation (A1) is cast in state-space form to backcast daily consumption data in the remote past. The equation is a univariate autoregressive model over weekly frequencies with a set of exogenous regressors composed of weekly seasonal log differences daily data of natural gas for thermoelectric use and by a year fixed seasonal cycle over a year (see Gómez [15]).

More in detail, D_t is obtained by the following equation according to Gregorian calendar with a period of 365.25 days:

$$s_t = a \cdot \cos(wt) + b \cdot \sin(wt),$$

$$w = 2\pi \cdot \frac{k}{n} \quad (\text{A2})$$

For a more complex example considering two fixed seasonal patterns (according to the Gregorian and the Hiri calendar), see Livera et al. [30]. For all the explanations concerning filtering and smoothing (used here to backcast data), see Chapter 4 of Gómez [14].

Appendix A.2. Model Synthesis

Table A1. Synthesis of state–space models used for energy.

Index	Log	Model	Step 1	Step 2	Step 3
IPI-EXTRA	YES	SUTSE	CH4NAZ	CH4NAZ25days	NaN
IPI-PETROL	NO	ARIMA	Automatic	Automatic	Automatic
IPI-PRODELE	YES	SUTSE	TERNAPROD	CONSUMPTION25days	NaN
IPI-CNG	YES	SUTSE	CH4TRANSPORTED	CH4TRANSPORTED25days	NaN

Table A2. Synthesis of state–space models used for energy at daily frequencies.

Index	Week	Model	Step 1	Step 2
CH4-PRODNAZ	Yes	ARIMA	25 days	NaN
CONSUMPTION	Yes	ARIMA	25 days	NaN
CH4-THERMO	Yes	ARIMA	25 days	NaN
CH4-TRANSPORTED	Yes	ARIMA	25 days	NaN

Table A3. Synthesis of Seasonal Arima used for energy at daily frequencies.

Index	No Log	p	d	q	ps	ds	qs
CH4-PRODNAZ	true	1	0	1	0	1	0
CONSUMPTION	true	3	0	0	0	1	0
CH4-TRANSPORTED	true	2	0	0	0	1	1

According to Gómez [15]:

- p: degree of regular AR polynomial
- d: degree of regular differencing
- q: degree of regular MA polynomial
- ps: degree of seasonal AR polynomial
- ds: degree of seasonal differencing
- qs: degree of seasonal MA polynomial.

Appendix B

Table A4. Barebone model variables: IPI_t intermediate goods $TTMI_t$ in Germany.

Parameter	Estimate	t-Ratio
$\sigma_{level11}$	1.1845	NaN
$\sigma_{level21}$	0.5328	7.5028
$\sigma_{level22}$	0.3730	5.3705
σ_{seas11}	0.0000	NaN
σ_{seas21}	0.2027	5.4352
σ_{seas22}	0.0000	NaN
$\sigma_{irreg11}$	1.2117	4.1055
$\sigma_{irreg21}$	0.8714	4.3232
$\sigma_{irreg22}$	1.0022	4.8066
$\beta_{LS0902ITA}$	1.06778	5.40
$\beta_{LS0902GER}$	0.30503	2.26
$\beta_{LS0812ITA}$	1.25554	7.12
$\beta_{LS0812GER}$	0.51971	4.36
$\beta_{LS0908ITA}$	1.05184	4.38
$\beta_{LS0908GER}$	0.10807	0.72
β_{WDITA1}	3.16489	7.78
β_{WDITA2}	0.16782	0.56
β_{WDGER1}	0.63938	1.72
β_{WDGER2}	3.15687	11.38
$\beta_{LTRENDITA}$	−0.10771	−1.28
$\beta_{LTRENDGERM}$	0.19798	4.25
$\beta_{LS0401ITA}$	0.33050	0.14
$\beta_{LS0401ITA}$	−0.62274	−0.43
σ_{ϵ_1}		3.3784
σ_{ϵ_2}		2.3052
Parameter $\sigma_{level11}$ is concentrated out of the likelihood		
AIC		4050.8877
BIC		4131.1130

Table A5. Barebone model variables: IPI_t intermediate goods $LCVDE_t$ in Germany.

Parameter	Estimate	t-Ratio
$\sigma_{level11}$	1.0361	NaN
$\sigma_{level21}$	0.2124	3.6125
$\sigma_{level22}$	0.3613	5.2296
σ_{seas11}	0.0000	NaN
σ_{seas21}	0.2261	5.8753
σ_{seas22}	0.0000	NaN
$\sigma_{irreg11}$	1.0233	3.2737
$\sigma_{irreg21}$	0.2602	1.7839
$\sigma_{irreg22}$	0.6933	4.8188
$\beta_{LS0902ITA}$	1.11999	6.05
$\beta_{LS0902GER}$	0.14078	1.59
$\beta_{LS0812ITA}$	1.28269	7.77
$\beta_{LS0812GER}$	0.22200	2.82
$\beta_{LS0908ITA}$	1.09914	4.96
$\beta_{LS0908GER}$	−0.02810	−0.28
$\beta_{AO064ITA}$	−0.33737	−1.39
$\beta_{AO064GER}$	1.01140	8.21
$\beta_{AO1612ITA}$	1.73391	3.28
$\beta_{AO1612GERM}$	1.04162	3.86
β_{WDITA1}	3.20568	8.00
β_{WDITA2}	−0.28366	−1.40
$\beta_{WDGERM1}$	0.69936	1.92
$\beta_{WDGERM2}$	1.18202	6.39
$\beta_{LTRENDITA}$	−0.10538	−1.45
$\beta_{LTRENDGERM}$	0.05833	1.97
σ_{ϵ_1}		3.3051
σ_{ϵ_2}		1.5841
Parameter $\sigma_{level11}$ is concentrated out of the likelihood		
AIC		3921.7404
BIC		4009.9883

Table A6. Barebone model variables: IPI_t intermediate goods $LCVFR_t$ in France.

Parameter	Estimate	t-Ratio
$\sigma_{level_{11}}$	1.1822	NaN
$\sigma_{level_{21}}$	0.3091	3.4874
$\sigma_{level_{22}}$	0.4283	4.8279
$\sigma_{seas_{11}}$	0.0000	NaN
$\sigma_{seas_{21}}$	0.2044	5.1963
$\sigma_{seas_{22}}$	0.0000	NaN
$\sigma_{irreg_{11}}$	1.1527	3.8582
$\sigma_{irreg_{21}}$	0.0000	NaN
$\sigma_{irreg_{22}}$	1.2818	6.2179
$\beta_{LS0902ITA}$	1.06634	5.40
$\beta_{LS0902FRA}$	−0.04654	−0.37
$\beta_{LS0812ITA}$	0.80955	3.85
$\beta_{LS0812FRA}$	0.31748	2.11
$\beta_{LS0908ITA}$	0.99775	4.18
$\beta_{LS0908FRA}$	0.10745	0.80
$\beta_{LS0712ITA}$	1.30024	3.89
$\beta_{LS0712FRA}$	0.48621	2.14
β_{WDITA1}	3.48145	8.90
β_{WDITA2}	0.39849	1.46
β_{WDFRA1}	0.27574	0.81
β_{WDFRA2}	0.91456	3.83
$\beta_{LTRENDITA}$	−0.08364	−1.01
$\beta_{LTRENDGERM}$	0.04172	1.11
σ_{ϵ_1}		3.4133
σ_{ϵ_2}		2.1920
Parameter $\sigma_{level_{11}}$ is concentrated out of the likelihood		
AIC		4075.3171
BIC		4151.5312

Table A7. Barebone model variables: IPI_t intermediate goods $LCVES_t$ in Spain.

Parameter	Estimate	t-Ratio
$\sigma_{level_{11}}$	1.1822	NaN
$\sigma_{level_{21}}$	0.3091	3.4874
$\sigma_{level_{22}}$	0.4283	4.8279
$\sigma_{seas_{11}}$	0.0000	NaN
$\sigma_{seas_{21}}$	0.2044	5.1963
$\sigma_{seas_{22}}$	0.0000	NaN
$\sigma_{irreg_{11}}$	1.1527	3.8582
$\sigma_{irreg_{21}}$	0.0000	NaN
$\sigma_{irreg_{22}}$	1.2818	6.2179
$\beta_{LS0902ITA}$	1.09510	5.62
$\beta_{LS0902SPA}$	0.14489	1.73
$\beta_{LS0812ITA}$	1.25567	7.21
$\beta_{LS0812SPAIN}$	0.16982	2.27
$\beta_{LS0908ITA}$	1.03560	4.46
$\beta_{LS0908SPAIN}$	0.22975	2.26
$\beta_{LS061ITA}$	−0.05388	−0.27
$\beta_{LS061SPAIN}$	0.98863	11.43
$\beta_{LS081ITA}$	−0.63779	−1.59
$\beta_{LS081SPAIN}$	1.12031	6.38
$\beta_{TC054ITA}$	−0.05465	−0.09
$\beta_{TC054SPAIN}$	1.01704	4.14
$\beta_{LS086ITA}$	1.67800	3.12
$\beta_{LS086SPAIN}$	1.09436	4.66
β_{WDITA1}	3.22807	9.08
β_{WDITA2}	0.48477	3.25
β_{WDFRA1}	0.60840	1.98
β_{WDFRA2}	0.27005	2.10
$\beta_{LTRENDITA}$	−0.09225	−1.19
$\beta_{LTRENDGERM}$	0.08284	2.31
σ_{ϵ_1}		3.4853
σ_{ϵ_2}		1.4908
Parameter $\sigma_{level_{11}}$ is concentrated out of the likelihood		
AIC		3908.4210
BIC		4016.7252

Table A8. Endogenous variables: IPI_t intermediate goods $TTMI_t$ in Germany.

Parameter	Estimate	t-Ratio
$\sigma_{level11}$	1.2754	NaN
$\sigma_{level21}$	0.4541	5.4054
$\sigma_{level22}$	0.3707	5.0575
σ_{seas11}	0.0048	−0.2245
σ_{seas21}	0.2097	5.5352
σ_{seas22}	0.0000	NaN
$\sigma_{irreg11}$	0.7965	3.4383
$\sigma_{irreg21}$	1.1091	5.9255
$\sigma_{irreg22}$	0.3019	0.9444
$\beta_{LS0902ITA}$	0.97539	5.10
$\beta_{LS0902GER}$	0.22781	1.84
$\beta_{LS0812ITA}$	0.96489	5.57
$\beta_{LS0812GER}$	0.36516	3.25
$\beta_{LS0908ITA}$	0.96095	4.05
$\beta_{LS0908GER}$	0.11530	0.85
β_{WDITA1}	−1.11149	−0.48
β_{WDITA2}	−0.87904	−0.66
β_{WDGER1}	3.32781	8.86
β_{WDGER2}	0.39080	1.41
$\beta_{LTRENDITA}$	−0.18370	−0.50
$\beta_{LTRENDGERM}$	2.51376	9.16
$\beta_{LCVITA1}$	−0.14214	−1.58
$\beta_{LCVITA2}$	0.16423	3.88
$\beta_{LCVGERM1}$	0.00039	4.62
$\beta_{LCVGERM2}$	0.00008	1.52
$\beta_{LS0412ITA}$	0.00040	3.38
$\beta_{LS0412GER}$	0.00055	6.35
σ_{ϵ_1}		3.2395
σ_{ϵ_2}		2.1093
Parameter $\sigma_{level11}$ is concentrated out of the likelihood		
AIC		4007.5646
BIC		4107.8463

Table A9. Endogenous variables: IPI_t capital goods $TTMI_t$ in Germany.

Parameter	Estimate	t-Ratio
$\sigma_{level11}$	1.0740	NaN
$\sigma_{level21}$	0.4663	7.9032
$\sigma_{level22}$	0.3079	4.3649
σ_{seas11}	0.0000	NaN
σ_{seas21}	0.1542	4.2180
σ_{seas22}	0.0000	NaN
$\sigma_{irreg11}$	1.9223	6.2424
$\sigma_{irreg21}$	0.5257	3.0287
$\sigma_{irreg22}$	1.3111	6.5791
$\beta_{LS0811ITA}$	0.64013	2.58
$\beta_{LS0811GERM}$	0.28580	1.77
$\beta_{LS0902ITA}$	0.78372	4.42
$\beta_{LS0902GERM}$	0.24890	2.29
$\beta_{LS0802ITA}$	0.87032	3.84
$\beta_{LS0802GERM}$	0.29600	2.19
$\beta_{LS0808ITA}$	0.85618	2.23
$\beta_{LS0808GERM}$	0.62724	2.49
β_{WDITA1}	3.69889	8.01
β_{WDITA2}	0.17520	0.58
β_{WDGER1}	0.15927	0.37
β_{WDGER2}	3.09418	10.92
$\beta_{LTRENDITA}$	0.11424	1.49
$\beta_{LTRENDGERM}$	0.22951	5.71
$\beta_{LS0412ITA}$	2.11634	0.89
$\beta_{LS0412GER}$	−0.52124	−0.38
σ_{ϵ_1}		3.5837
σ_{ϵ_2}		2.2536
Parameter $\sigma_{level11}$ is concentrated out of the likelihood		
AIC		4078.9627
BIC		4151.1655

Table A10. Endogenous variables: IPI_t capital goods $TTMI_t$ in Germany.

Parameter	Estimate	t-Ratio
$\sigma_{level11}$	1.0984	NaN
$\sigma_{level21}$	0.5051	7.1653
$\sigma_{level22}$	0.3435	5.2124
σ_{seas11}	0.0034	0.1592
σ_{seas21}	−0.1124	−2.9125
σ_{seas22}	0.0000	NaN
$\sigma_{irreg11}$	1.4849	4.9785
$\sigma_{irreg21}$	0.2746	1.6744
$\sigma_{irreg22}$	0.9444	5.6833
$\beta_{LS0811ITA}$	0.21208	1.24
$\beta_{LS0811GERM}$	0.21766	1.95
$\beta_{LS0902ITA}$	0.62362	4.51
$\beta_{LS0902GERM}$	0.22689	2.59
$\beta_{LS0808ITA}$	0.27547	1.42
$\beta_{LS0808GERM}$	0.12754	1.03
β_{WDITA1}	−0.77571	−0.38
β_{WDITA2}	−0.51911	−0.42
β_{WDGER1}	2.95664	8.05
β_{WDGER2}	0.30505	1.21
$\beta_{LTRENDITA}$	−0.34809	−0.95
$\beta_{LTRENDGERM}$	2.11117	8.46
$\beta_{LCVITA1}$	0.05194	0.72
$\beta_{LCVITA2}$	0.20176	4.85
$\beta_{LCVGERM1}$	0.32615	11.17
$\beta_{LCVGERM2}$	0.07065	3.70
$\beta_{LS0412ITA}$	0.08949	1.58
$\beta_{LS0412GER}$	0.32594	8.55
σ_{ϵ_1}		2.8376
σ_{ϵ_2}		1.8952
Parameter $\sigma_{level11}$ is concentrated out of the likelihood		
AIC		3924.8623
BIC		4005.0876

Table A11. Endogenous variables: IPI_t capital goods $LCVDE_t$ in Germany.

Parameter	Estimate	t-Ratio
$\sigma_{level11}$	0.9588	NaN
$\sigma_{level21}$	0.2376	4.6892
$\sigma_{level22}$	0.2919	4.1896
σ_{seas11}	0.0189	1.1352
σ_{seas21}	0.1833	4.8999
σ_{seas22}	0.0000	NaN
$\sigma_{irreg11}$	1.9132	6.1456
$\sigma_{irreg21}$	0.0064	0.0512
$\sigma_{irreg22}$	0.7469	5.3079
$\beta_{LS0811ITA}$	0.93367	4.77
$\beta_{LS0811GER}$	0.28173	3.57
$\beta_{LS0902ITA}$	0.93005	5.73
$\beta_{LS0902GER}$	0.15037	2.31
$\beta_{LS0808ITA}$	0.92789	4.22
$\beta_{LS0808GER}$	0.03098	0.35
$\beta_{AO0604ITA}$	−0.33339	−1.14
$\beta_{AO0604GER}$	0.98745	8.25
$\beta_{AO0612ITA}$	0.98016	1.53
$\beta_{AO0612GER}$	1.08035	4.13
β_{WDITA1}	3.81575	8.05
β_{WDITA2}	−0.29295	−1.50
β_{WDGER1}	0.06230	0.14
β_{WDGER2}	1.14210	6.37
$\beta_{LTRENDITA}$	0.10298	1.52
$\beta_{LTRENDGERM}$	0.06338	2.38
σ_{ϵ_1}		3.6460
σ_{ϵ_2}		1.5178
Parameter $\sigma_{level11}$ is concentrated out of the likelihood		
AIC		3954.0442
BIC		4046.3033

Table A12. Endogenous variables: IPI_t capital goods $LCVFR_t$ in France.

Parameter	Estimate	t-Ratio
$\sigma_{level11}$	1.1694	NaN
$\sigma_{level21}$	0.3471	4.6773
$\sigma_{level22}$	0.3395	4.2102
σ_{seas11}	0.0000	NaN
σ_{seas21}	0.1624	4.0324
σ_{seas22}	0.0000	NaN
$\sigma_{irreg11}$	1.8498	6.0017
$\sigma_{irreg21}$	0.0073	0.0431
$\sigma_{irreg22}$	1.2797	6.2774
$\beta_{LS0811ITA}$	0.60877	2.43
$\beta_{LS0811FRA}$	0.37689	2.48
$\beta_{LS0902ITA}$	0.79028	4.33
$\beta_{LS0902FRA}$	−0.06272	−0.62
$\beta_{LS0808ITA}$	0.91365	3.90
$\beta_{LS0808FRA}$	0.07199	0.58
$\beta_{LS0812ITA}$	0.93976	2.30
$\beta_{LS0812FRA}$	0.52237	2.08
β_{WDITA1}	3.68414	8.37
β_{WDITA2}	0.36392	1.34
β_{WDFRA1}	0.20422	0.53
β_{WDFRA2}	0.93457	3.95
$\beta_{LTRENDITA}$	0.10829	1.32
$\beta_{LTRENDFRA}$	0.05151	1.49
σ_{ϵ_1}		3.6843
σ_{ϵ_2}		2.1486
Parameter $\sigma_{level11}$ is concentrated out of the likelihood		
AIC		4103.5609
BIC		4183.7862

Table A13. Endogenous variables: IPI_t capital goods $LCVES_t$ in Spain.

Parameter	Estimate	t-Ratio
$\sigma_{level11}$	2.0010	NaN
$\sigma_{level21}$	0.2001	40032.8244
$\sigma_{level22}$	0.2000	2122.4735
σ_{seas11}	0.2002	3128.5844
σ_{seas21}	0.2001	4228.3941
σ_{seas22}	0.0000	NaN
$\sigma_{irreg11}$	0.2001	28197.7772
$\sigma_{irreg21}$	0.2002	1871.4497
$\sigma_{irreg22}$	0.2001	2797.4601
$\beta_{LS0811ITA}$	0.72268	3.44
$\beta_{LS0811SPA}$	0.22154	2.09
$\beta_{LS0902ITA}$	0.54647	2.75
$\beta_{LS0902SPA}$	0.03201	0.40
$\beta_{LS0808ITA}$	0.92888	3.36
$\beta_{LS0808SPA}$	0.10834	0.71
$\beta_{LS0812ITA}$	−0.44912	−1.83
$\beta_{LS0812SPA}$	0.88650	12.98
$\beta_{LS0601ITA}$	−1.63635	−3.27
$\beta_{LS0601SPA}$	1.03738	6.12
$\beta_{LS0801ITA}$	1.00302	1.53
$\beta_{LS0801SPA}$	1.02250	3.09
$\beta_{LS0806ITA}$	0.22150	3.44
$\beta_{LS0806SPA}$	0.06790	2.09
β_{WDITA1}	3.60571	10.12
β_{WDITA2}	0.57797	2.04
β_{WDSPA1}	0.24221	0.81
β_{WDSPA2}	0.27781	1.16
$\beta_{LTRENDITA}$	0.00694	0.05
$\beta_{LTRENDSPA}$	0.09744	4.63
σ_{ϵ_1}		3.9909
σ_{ϵ_2}		2.6079
Parameter $\sigma_{level11}$ is concentrated out of the likelihood		
AIC		4167.8550
BIC		4276.1592

Table A14. Endogenous variables: IPI_t durable goods $TTMI_t$ in Germany.

Parameter	Estimate	t-Ratio
$\sigma_{level_{11}}$	1.0978	NaN
$\sigma_{level_{21}}$	0.4715	4.3692
$\sigma_{level_{22}}$	0.4598	4.5298
$\sigma_{seas_{11}}$	0.0035	0.1760
$\sigma_{seas_{21}}$	0.3480	5.1223
$\sigma_{seas_{22}}$	0.0000	NaN
$\sigma_{irreg_{11}}$	3.4814	5.1377
$\sigma_{irreg_{21}}$	0.1092	0.5888
$\sigma_{irreg_{22}}$	1.3653	5.3712
$\beta_{LS0801ITA}$	−16.83680	−5.51
$\beta_{LS0801GER}$	−9.64008	−6.39
β_{WDITA1}	3.21904	3.92
β_{WDITA2}	−0.00274	−0.01
β_{WDGER1}	1.67770	2.21
β_{WDGER2}	3.31421	11.77
$\beta_{LTRENDITA}$	−0.21005	−2.66
$\beta_{LTRENDGERM}$	0.19630	4.22
$\beta_{LS01010412ITA}$	−1.15132	−0.38
$\beta_{LS01010412GER}$	−0.69390	−0.46
σ_{ϵ_1}		6.0879
σ_{ϵ_2}		2.3218
Parameter $\sigma_{level_{11}}$ is concentrated out of the likelihood		
AIC		4331.5822
BIC		4379.7174

Table A15. Endogenous variables: IPI_t non-durable goods, CNG_t Industrial.

Parameter	Estimate	t-Ratio
$\sigma_{level_{11}}$	0.7388	NaN
$\sigma_{level_{21}}$	0.4359	0.6489
$\sigma_{level_{22}}$	2.4961	3.6884
$\sigma_{seas_{11}}$	0.1133	2.2909
$\sigma_{seas_{21}}$	0.0994	1.9265
$\sigma_{seas_{22}}$	0.0790	0.9532
$\sigma_{irreg_{11}}$	1.9753	3.2214
$\sigma_{irreg_{21}}$	0.2920	0.8076
$\sigma_{irreg_{22}}$	1.4162	2.1224
β_{WDNDUR}	0.14242	0.98
β_{WDCNG}	0.96736	5.59
$\beta_{AO0412NDUR}$	0.13234	0.73
$\beta_{AO0412CNG}$	0.92258	4.37
$\beta_{AO0602NDUR}$	3.13904	12.63
$\beta_{AO0602CNG}$	0.24930	0.88
$\beta_{CONSELENDUR}$	0.00021	2.08
$\beta_{CONSELECNG}$	−0.00012	−0.87
$\beta_{LCVITNDUR}$	0.00066	1.75
$\beta_{LCVITCNG}$	0.00257	5.29
σ_{ϵ_1}		3.0291
σ_{ϵ_2}		4.1020
Parameter $\sigma_{level_{11}}$ is concentrated out of the likelihood		
AIC		3474.8432
BIC		3543.5512

Table A16. Endogenous variables: log production of electricity log cumulated daily consumption of electricity.

Parameter	Estimate	t-Ratio
$\sigma_{level_{11}}$	0.0149	NaN
$\sigma_{level_{21}}$	0.0038	3.3241
$\sigma_{level_{22}}$	0.0044	5.2690
$\sigma_{seas_{11}}$	0.0001	−0.1268
$\sigma_{seas_{21}}$	0.0001	0.1369
$\sigma_{seas_{22}}$	0.0000	0.0907
$\sigma_{irreg_{11}}$	0.0206	6.4089
$\sigma_{irreg_{21}}$	0.0144	7.0658
$\sigma_{irreg_{22}}$	0.0100	3.9126
σ_{ϵ_1}		0.0370
σ_{ϵ_2}		0.0278
Parameter $\sigma_{level_{11}}$ is concentrated out of the likelihood		
BIC		376.56
AIK		384.58

Table A17. Endogenous variables: log IPI extraction of crude petroleum and natural gas log production of natural gas.

Parameter	Estimate	t-Ratio
$\sigma_{level_{11}}$	0.0270	NaN
$\sigma_{level_{21}}$	0.0171	13.7116
$\sigma_{level_{22}}$	0.0079	7.3941
$\sigma_{seas_{11}}$	0.0014	6.1766
$\sigma_{seas_{21}}$	0.0043	5.6644
$\sigma_{seas_{22}}$	0.0001	−0.1410
$\sigma_{irreg_{11}}$	0.0242	9.3619
$\sigma_{irreg_{21}}$	0.0145	4.0573
$\sigma_{irreg_{22}}$	0.0167	3.9616
σ_{ϵ_1}		0.0727
σ_{ϵ_2}		0.0515
Parameter $\sigma_{level_{11}}$ is concentrated out of the likelihood		
BIC		867.38
AIK		875.40

Table A18. Endogenous variables: log IPI manufacture of gas; distribution of gaseous fuels through mains log natural gas transported.

Parameter	Estimate	t-Ratio
$\sigma_{level_{11}}$	0.0378	NaN
$\sigma_{level_{21}}$	0.0341	18.8915
$\sigma_{level_{22}}$	0.0057	5.4502
$\sigma_{seas_{11}}$	0.0038	NaN
$\sigma_{seas_{21}}$	0.0054	12.057
$\sigma_{seas_{22}}$	0.0038	NaN
$\sigma_{irreg_{11}}$	0.0038	NaN
$\sigma_{irreg_{21}}$	0.0013	−0.3442
$\sigma_{irreg_{22}}$	0.0038	NaN
σ_{ϵ_1}		0.0766
σ_{ϵ_2}		0.0724
Parameter $\sigma_{level_{11}}$ is concentrated out of the likelihood		
BIC		872.70
AIK		880.72

Table A19. Long weekend model dependent variable: IPI $(0,0,1)(0,1,1)_{12}$ *nolog*.

Parameter	Estimate	t-Ratio
θ_1	0.7559	18.3690
θ_{12}	−0.1882	−2.2772
β_{WD}	3.40012	13.26
$\beta_{NEWYEAR}$	0.38226	0.31
$\beta_{EPIPHANY}$	−2.34537	−1.44
$\beta_{REPUBLICDAY}$	−0.75782	−0.56
$\beta_{IMMACULATE}$	−1.09217	−0.76
$\beta_{WORKERSDAY}$	1.09799	0.90
Parameter σ_ϵ is concentrated out of the likelihood		
BIC		2127.4359
AIK		2153.9809

Table A20. Dependent variable: IPI manufacture of coke and refined petroleum products $(0,1,1)(0,1,1)_{12}$ *nolog*.

Parameter	Estimate	t-Ratio
θ_1	−0.8844	−28.48
θ_{12}	−0.4798	−6.16
β_{AO0605}	−22.81603	−5.61
β_{AO0803}	20.07762	4.95
β_{AO0411}	−15.59829	−3.81
β_{AO0212}	−15.02531	−3.63
β_{TC0203}	15.42229	6.32
β_{AO1006}	−18.32463	−7.65
β_{AO0711}	−14.53021	−7.18
β_{LS0504}	8.61470	3.64
β_{LS1211}	−15.34803	−6.45
β_{LS1502}	10.36593	7.74
β_{TC0911}	−12.05044	−8.93
β_{TC0911}	13.36453	10.35
β_{TC0311}	−11.53890	−5.96
β_{AO1611}	−7.45864	−3.69
β_{TC1802}	−10.33852	−5.53
β_{TC0106}	9.70640	4.11
β_{TC0102}	10.45859	4.57
β_{TC1706}	−11.93867	−5.11
β_{AO0405}	−9.59695	−3.63
β_{TC0511}	6.21901	3.05
β_{TC1111}	8.58217	3.63
β_{LS1002}	−7.92531	−4.04
β_{LS0805}	−9.49671	−4.82
β_{LS1701}	6.98222	5.17
β_{TC0702}	−7.00933	−5.12
β_{AO1009}	6.31182	4.26
β_{AO0904}	6.02294	3.21
β_{AO1505}	−7.31103	−3.12
β_{AO1206}	7.94524	3.35
β_{AO0710}	6.80889	2.89
β_{AO0410}	−7.37313	−3.12
β_{AO0402}	−7.31373	−3.10
β	−6.83603	−2.89
β	6.11984	2.64
σ_ϵ		2.7709
BIC		1878.96
AIK		1872.34

Appendix B.1. Seasonal VARMA Model

Bruno et al. [4] used the following unrestricted vector autoregressive model:

$$\Delta\Delta_{12}y_t = \beta\Delta_{12}y_{t-1} + \sum_{j=1}^{13}\gamma_j\Delta\Delta_{12}y_{t-j} + \phi d_t + \epsilon_t \quad (\text{A3})$$

where $\Delta = (1 - L)$, $\Delta_{12} = (1 - L^{12})$ and L is the usual lag operator such that $L^p z_t = z_{t-p}$, $y_t = (\log(IIP_t); \log(TONFSt); PP_t)$. (PP_t represents production prospects from ISTAT surveys, and $TONFSt$ stands for tons of raw material transported by Italian railways. Bruno and Lupi [4] applied to PP_t a logistic transformation, d_t contains some deterministic components (constant, specific impulse dummies.) The endogenous variables available at the time $t + 1$ of forecast are $TONFS_t$ and PP_t . For the next periods, the forecast is unconditional since there is no further information. Since intermediate goods plus capital goods together account for 61 per cent of the whole index and they are transported on light commercial vehicles, I take the sum of their registration over Italy, France, Spain and Germany. Thus, the new model is

$$\Delta\Delta_{12}y_t = \phi d_t + \gamma\Delta\Delta_{12}y_{t-1} + \theta_k \epsilon_{t-1} \quad (\text{A4})$$

where $\Delta = (1 - L)$, $\Delta_{12} = (1 - L^{12})$ and L is the usual lag operator such that $L^p z_t = z_{t-p}$, $y_t = (\log(IPIINTERMEDIATE_t); \log(IPICAPITAL_t); \log(LCVSUM_t))$ ($LCVSUM_t$ represents the total level of registrations of light commercial vehicles over the four mentioned European countries, while d_t contains the long weekend dummies, the growth rates of number of days worked over the four countries and the growth rates of number of days worked in Italy computed as first differences as first log year to year differences) (γ and θ for simplicity have no further lags to ease estimation with Hannan–Rissanen method, see Gómez [15]).

It is also possible to compute a bivariate VARMA system to evaluate the importance of single country LCV registration over Italian Production Index of Intermediate Goods and over Italian Production Index of Capital Goods.

B.2. Performance of Conditional VARMA over Three Steps

Table A21 summarizes the performance of the conditional vector autoregressive model using different light van registration and the truck index. From a comparison of Tables A21–A28, it is possible to check how the barebone model beats conditional VARMA over three steps. Similarly, by means of Model Confidence Set Procedure (see Tables A25–A28), we show the supremacy of the enhanced model with respect to the conditional VARMA alternatives.

At Step 1, the system using the truck index as endogenous variable available at time $t + 1$ always wins the competition. This advantage is not so marked at Steps 2 and 3.

Table A21. Comparison of mean absolute error with expanding window applied to conditional var with three variables. The forecast windows expands itself from 2008:01 to 2018:12. The ratio between MAEs is shown. A value less than one shows an improvement over benchmark.

	Step 1		Step 2		Step 3	
	IPI-INT	IPI-CAP	IPI-INT	IPI-CAP	IPI-INT	IPI-CAP
LCV-SUM	0.5392	0.5648	0.6994	0.7289	0.7074	0.7619
LCV-ITA	0.8725	0.8992	0.7385	0.7344	0.8032	0.8544
LCV-GER	0.6644	0.6834	0.7242	0.7614	0.7340	0.7729
LCV-FRA	0.8217	0.8025	0.7732	0.7951	0.7877	0.8703
LCV-SPA	0.7332	0.7723	0.7110	0.7642	0.7420	0.7858
TTMI	0.4896	0.4852	0.7043	0.7821	0.7082	0.8018

Table A22. Comparison of root mean squared error with expanding window applied to conditional var with three variables. The forecast windows expands itself from 2008:01 to 2018:12. The ratio between RMSEs is showed. A value less than one shows an improvement over benchmark.

	Step 1		Step 2		Step 3	
	IPI-INT	IPI-CAP	IPI-INT	IPI-CAP	IPI-INT	IPI-CAP
LCV-SUM	0.4862	0.5687	0.6265	0.7250	0.6621	0.7711
LCV-ITA	1.1405	1.4685	0.7189	0.7313	1.0082	1.1193
LCV-GER	0.5803	0.6675	0.6386	0.7336	0.6987	0.7902
LCV-FRA	0.8912	0.9998	0.7483	0.8452	0.8403	0.9994
LCV-SPA	0.7010	0.8923	0.6656	0.7758	0.7339	0.8508
TTMI	0.4216	0.4674	0.6165	0.7490	0.6535	0.7932

Table A23. Comparison of mean absolute error with expanding window applied to conditional var with three variables. The forecast windows expands itself from 2015:01 to 2018:12. The ratio between MAEs is shown. A value less than one shows an improvement over benchmark.

	Step 1		Step 2		Step 3	
	IPI-INT	IPI-CAP	IPI-INT	IPI-CAP	IPI-INT	IPI-CAP
LCV-SUM	0.7694	0.5258	1.0942	0.7630	0.9811	0.6883
LCV-ITA	1.1604	0.7950	1.1291	0.7167	0.9077	0.6726
LCV-GER	0.9548	0.6766	1.0178	0.8397	0.9458	0.7705
LCV-FRA	1.0818	0.7495	1.1096	0.7672	0.8648	0.6616
LCV-SPA	1.0210	0.6814	1.0239	0.7269	0.8363	0.5922
TTMI	0.7428	0.5384	1.0637	0.8261	0.9844	0.7180

Table A24. Comparison of root mean squared error with expanding window applied to conditional var with three variables. The forecast windows expands itself from 2015:01 to 2018:12. The ratio between RMSEs is shown. A value less than one shows an improvement over benchmark.

	Step 1		Step 2		Step 3	
	IPI-INT	IPI-CAP	IPI-INT	IPI-CAP	IPI-INT	IPI-CAP
LCV-SUM	0.7838	0.5489	1.0692	0.7853	0.9875	0.6902
LCV-GER	0.9195	0.6840	0.9842	0.8748	0.9213	0.8064
LCV-ITA	1.1272	0.8163	1.0679	0.7395	0.8667	0.6506
LCV-FRA	1.0477	0.7754	1.0505	0.7771	0.8869	0.6629
LCV-SPA	0.9834	0.7256	0.9956	0.7552	0.8533	0.6180
TTMI	0.7313	0.5721	1.0384	0.8308	0.9445	0.7178

Table A25. Model confidence set hierarchy loss function is absolute error at the end of the sample for intermediate goods model comparison $\alpha = 0.05$, number of bootstrap replications = 5000 and block length = 12. The null hypothesis is that the average performance of the model in the row is as small as the minimum average performance across the remaining models. The alternative is that the minimum average loss across the remaining models is smaller than the average performance of the model in the row.

Index	Step 1	Step 2	Step 3
IPI-BEN-INTERM	0.0000	0.0020	0.0290
VARMATTMI	0.0000	0.0020	0.0320
VARMALCVGER	0.0000	0.0060	0.0320
VARMALCVITA	0.0000	0.0060	0.0320
VARMALCVFRA	0.0030	0.0060	0.0320
VARMALCVSPA	0.1450	0.0060	0.0380
VARMALCVSUM	0.2470	0.0060	0.0380
ENHMTTMI	1.0000	1.0000	1.0000

Table A26. Model confidence set hierarchy Loss function is squared error at the end of the sample for intermediate goods model comparison $\alpha = 0.05$, number of bootstrap replications = 5000 and block length = 12. The null hypothesis is that the average performance of the model in the row is as small as the minimum average performance across the remaining models. The alternative is that the minimum average loss across the remaining models is smaller than the average performance of the model in the row.

Index	Step 1	Step 2	Step 3
IPI-BEN-INTERM	0.0770	0.3160	0.2470
VARMMATTMI	0.0770	0.3160	0.2470
VARMALCVGER	0.0770	0.4110	0.6180
VARMALCVITA	0.0770	0.4110	0.6180
VARMALCVFRA	0.0770	0.4110	0.6180
VARMALCVSPA	0.0770	0.4110	0.8620
VARMALCVSUM	0.0770	0.4110	0.8620
ENHMTTMI	1.0000	1.0000	1.0000

Table A27. Model confidence set hierarchy loss function is absolute error at the end of the sample for capital goods model comparison $\alpha = 0.05$, number of bootstrap replications = 5000, and block length = 12. The null hypothesis is that the average performance of the model in the row is as small as the minimum average performance across the remaining models. The alternative is that the minimum average loss across the remaining models is smaller than the average performance of the model in the row.

Index	Step 1	Step 2	Step 3
IPI-BEN-CAPITAL	0.0000	0.0000	0.0000
VARMMATTMI	0.0000	0.0100	0.0110
VARMALCVGER	0.0000	0.0210	0.0510
VARMALCVITA	0.0000	0.0210	0.0510
VARMALCVFRA	0.0000	0.0210	0.0510
VARMALCVSPA	0.0040	0.0210	0.0510
VARMALCVSUM	0.1350	0.0290	0.0510
ENHMTTMI	1.0000	1.0000	1.0000

Table A28. Model confidence set hierarchy loss function is squared error at the end of the sample for capital goods model comparison $\alpha = 0.05$, number of bootstrap replications = 5000 and block length = 12. The null hypothesis is that the average performance of the model in the row is as small as the minimum average performance across the remaining models. The alternative is that the minimum average loss across the remaining models is smaller than the average performance of the model in the row.

Index	Step 1	Step 2	Step 3
IPI-BEN-CAPITAL	0.0000	0.0000	0.0000
VARMMATTMI	0.0110	0.1640	0.1630
VARMALCVGER	0.0110	0.1640	0.1630
VARMALCVITA	0.0110	0.1640	0.1630
VARMALCVFRA	0.0110	0.1640	0.1630
VARMALCVSPA	0.0160	0.1640	0.8640
VARMALCVSUM	0.5980	0.8460	0.8640
ENHMTTMI	1.0000	1.0000	1.0000

References

1. Bulligan, G.; Golinelli, R.; Parigi, G. Forecasting monthly industrial production in real-time: from single equations to factor-based models. *Empir. Econ.* **2010**, *39*, 303–336. doi:10.1007/s00181-009-0305-7. [CrossRef]
2. Marchetti, D.; Parigi, G. Energy consumption, survey data and the prediction of industrial production in Italy: A comparison and combination of different models. *J. Forecast.* **2000**, *19*, 419–440. doi:10.1002/1099-131X(200009)19:5<419::AID-FOR749>3.0.CO;2-J. [CrossRef]
3. Bodo, G.; Signorini, L.F. Short-term forecasting of the industrial production index. *Int. J. Forecast.* **1987**, *3*, 245–259. doi:10.1016/0169-2070(87)90006-9. [CrossRef]
4. Bruno, G.; Lupi, C. Forecasting industrial production and the early detection of turning points. *Empir. Econ.* **2004**, *29*, 647–671. [CrossRef]
5. Costantini, M. Forecasting the industrial production using alternative factor models and business survey data. *J. Appl. Stat.* **2013**, *40*, 2275–2289. doi:10.1080/02664763.2013.809870. [CrossRef]
6. Brunhes-Lesage, V.; Darné, O. Nowcasting the French index of industrial production: A comparison from bridge and factor models. *Econ. Model.* **2012**, *29*, 2174–2182. doi:10.1016/j.econmod.2012.04.011. [CrossRef]
7. Askita, N.; Zimmermann, K.F. Nowcasting Business Cycles Using Toll Data. *J. Forecast.* **2013**, *32*, 299–306. [CrossRef]
8. Cox, M.; Berghausen, M.; Linz, S.; Fries, C.; Völker, J. Digitale Prozessdaten aus der Lkw-Mauterhebung neuer Baustein der amtlichen Konjunkturstatistiken. *WISTA Wirtschaft und Statistik* **2018**, *2018*, 11.
9. Hansen, P.R.; Lunde, A.; Nason, J.M. The Model Confidence Set. *Econometrica* **2011**, *79*, 453–497. [CrossRef]
10. Durbin, J.; Koopman, S. *Time Series Analysis by State Space Methods*; Oxford University Press: Oxford, UK, 2012.
11. Gómez, V.; Maravall, A. Automatic Modeling Methods for Univariate Series. In *A Course in Time Series Analysis*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2000.
12. Gómez, V. *Linear Time Series with MATLAB and OCTAVE*; Springer: Berlin, Germany, 2019.
13. Harvey, A. *Forecasting, Structural Time Series Models and the Kalman Filter*; Cambridge University Press: Cambridge, UK, 1989.
14. Gómez, V. *Multivariate Time Series with Linear State Space Structure*; Springer: Berlin, Germany, 2016. doi:10.1007/978-3-319-28599-3. [CrossRef]
15. Gómez, V. SSMATLAB: A Set of MATLAB Programs for the Statistical Analysis of State Space Models. *J. Stat. Softw.* **2015**, *66*, 1–37. doi:10.18637/jss.v066.i09. [CrossRef]
16. Harvey, D.; Leybourne, S.; Newbold, P. Testing the equality of prediction mean squared errors. *J. Bus. Econ. Stat.* **1997**, *13*, 281–291. [CrossRef]
17. Reality Check Tests. Available online: <https://github.com/PedroBSB/mlRFinance/wiki/Reality-Check-Tests> (accessed on 23 October 2019).
18. Hansen, P.R. A Test for Superior Predictive Ability. *J. Bus. Econ. Stat.* **2005**, *23*, 365–380. [CrossRef]
19. White, H. A Reality Check for Data Snooping. *Econometrica* **2000**, *68*, 1097–1126. [CrossRef]
20. Bernardi, M.; Catania, L. The Model Confidence Set package for R. *arXiv* **2014**, arXiv:1410.8504.
21. Sheppard, K. MFE MATLAB Function Reference Financial Econometrics. Available online: <https://www.kevinshppard.com/files/code/matlab/mfe-toolbox-documentation.pdf> (accessed on 23 October 2019).
22. Hannan, E.J.; Rissanen, J. Recursive Estimation of Mixed Autoregressive-Moving Average Order. *Biometrika* **1982**, *69*, 81–94. [CrossRef]
23. Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*; Springer and Verlag: Berlin, Germany, 2005.
24. Reinsel, G.C. *Elements of Multivariate Time Series Analysis*; Springer and Verlag: New York, NY, USA, 1997.
25. Girardi, A.; Guardabascio, B.; Ventura, M. Factor-Augmented Bridge Models (FABM) and Soft Indicators to Forecast Italian Industrial Production. *J. Forecast.* **2016**, *35*, 542–552. [CrossRef]
26. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
27. Bowman, K.; Shenton, B. Omnibus test countours for departures from normality based on $\sqrt{b_1}$, and $\sqrt{b_2}$. *Biometrika* **1975**, *62*, 243–250.
28. Durbin, J.; Watson, G.S. Testing for Serial Correlation in Least Squares Regression: I. *Biometrika* **1950**, *37*, 409–428.

29. Durbin, J.; Watson, G.S. Testing for Serial Correlation in Least Squares Regression: II. *Biometrika* **1951**, *38*, 159–179. [[CrossRef](#)]
30. Livera, A.M.D.; Hyndman, R.J.; Snyder, R.D. Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *J. Am. Stat. Assoc.* **2011**, *106*, 1513–1527. doi:10.1198/jasa.2011.tm09771. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).