

*Błażej Mazur*\*

## Density Forecasts Based on Disaggregate Data: Nowcasting Polish Inflation\*\*

**A b s t r a c t.** The paper investigates gains in performance of density forecasts from models using disaggregate data when forecasting aggregate series. The problem is considered within a restricted VAR framework with alternative sets of exclusion restrictions. Empirical analysis of Polish CPI m-o-m inflation rate (using its 14 sub-categories for disaggregate modelling) is presented. Exclusion restrictions are shown to improve density forecasting performance (as evaluated using log-score and CRPS criteria) relatively to aggregate and also disaggregate unrestricted models.

**K e y w o r d s:** prediction, model comparison, density forecasting, inflation, VAR models, shrinkage.

**J E L Classification:** E31, E37, C53, C32.

### Introduction

The paper focuses on the question whether point and in particular density forecasts of univariate series can be improved using disaggregate information (assuming that it is available) – or more generally, whether economic fluctuations are more accurately modeled at the aggregate or at the disaggregate level. The crucial assumption being relevant here is that a multivariate model for disaggregate data is used only as a tool for obtaining the implied univariate forecast of the aggregate series. The aggregating weights are as-

---

\* Correspondence to: Błażej Mazur, Cracow University of Economics, Chair of Econometrics and Operations Research, e-mail: [eomazur@cyfronet.pl](mailto:eomazur@cyfronet.pl).

\*\* This research was supported by the Polish National Science Center (NCN) based on decision number DEC-2013/09/B/HS4/01945.

sumed to be known and fixed. There is no direct and necessary relationship between forecasting power in the disaggregate (multivariate) context and the validity of the implied univariate (aggregate) forecast. Therefore the usual statistical procedures used for forecast evaluation or model comparison applied at the disaggregate level have to be adjusted in order to take the context into account.

The probabilistic approach to inference have been gaining popularity in the systematic manner over time. In particular events like the recent economic crisis have made it evident that uncertainty quantification is an inherent part of the forecasting process. Some problems that seem to be uncomplicated when point forecast perspective is taken become more complex when one has the density forecasting perspective in mind. The paper investigates consequences of the aggregation for uncertainty of the aggregate forecast. However, an alternative perspective on the problems is the general question whether economic fluctuations are better modeled at the aggregate or the disaggregate level, given some specific model classes.

A framework for formal investigation of the problems mentioned above builds on a Gaussian VAR specification with restrictions. Usage of such relatively simple models is dictated by complexity that increases rapidly with the dimension (i.e. disaggregation level). Moreover, as the problem considered here can be interpreted in the context of variable selection (resulting in a very large number of possible exclusion restrictions), numerical complexity becomes quite considerable even within the relatively simple model class consisting of Gaussian VARs.

The objective of the paper is to investigate empirical evidence supporting the use of disaggregate data for aggregate forecasting. The application under consideration deals with forecasting Polish monthly CPI inflation rate (relative to the previous month) using 164 observations on growth rates of 14 price sub-indices. A pseudo real-time forecasting experiment is performed and out-of-sample predictive performance of competing specifications is evaluated over a verification window consisting of the last 44 observations. All the criteria (those relevant for point forecast and density forecasts) show the evidence of substantial gains from considering a large menu of competing specifications at the disaggregate level. However, as the total number of alternative restricted specifications is enormous, a stochastic search algorithm was used to explore the model space.

The results suggest that the problem of aggregate vs. disaggregate modeling for the sake of aggregate forecasting is not trivial and illustrates a potential for gains in predictive performance, but this requires some form of promoting parametric parsimony – as attributing non-zero probability to

models with zero restrictions might be interpreted as introducing some form of shrinkage within the overall (unrestricted) model.

The rest of the paper is organized in the following way: firstly, the general problem of forecasting of the aggregate series is presented in a more detailed way. Secondly, a formal model framework used here is described, with the relevant forecasting methodology. Thirdly, methods for evaluation of density forecasts are outlined. Fourthly, the empirical illustration is provided. A summary and some remarks about possible directions for further developments conclude.

## 1. Aggregate vs. Disaggregate Approach to Forecasting

A natural approach to forecasting is to use a model formulated directly in terms of the quantity of interest itself. However, if one considers the context of e.g. inflation forecasting, an approach that is often taken by practitioners is to obtain the aggregate forecast using individual disaggregate inflation forecasts for a number of sub-categories.

The aggregate approach seems to be conceptually simpler and more elegant. Moreover, due to the lack of the dimensionality problem, at the aggregate level it is possible to make use of more sophisticated models. Aggregation could have a regularizing impact as well. For example, one could expect that in certain cases the assumption of say homoscedastic Gaussian errors is more likely to hold at the aggregate level.

On the contrary, at the disaggregate level one has to deal with increased heterogeneity. Another issue is that of modeling dependencies across the variables in a multivariate process which might be challenging. However, there is more information available, and sometimes an expert knowledge can be utilized at the disaggregate level only.

In the specific context considered here, the only objective of the analysis is the aggregate forecast. It is not obvious that the efforts made to deal with heterogeneity and dependence modeling that improve goodness of fit at the disaggregate level would lead to an improved forecasting performance of the aggregate series. This is even more true when one has in mind the density forecasting context.

In general it is possible to obtain a point forecast of the aggregate series using disaggregate point forecasts obtained from individual, unrelated univariate models for the sub-aggregates. However, if one makes an attempt to generate a density forecast in the same way, the omitted dependence between sub-aggregates might result in uncertainty misspecification for the aggregate forecast, resulting in poor density predictive performance. The

tradeoffs between various aspects of heterogeneity and dependence modeling and possible gains in the aggregate density forecasting performance are among the points of interest in the paper. The essential empirical issue under consideration is whether there exist potential gains from working at the disaggregate level in terms of the aggregate density predictive performance. Moreover, one might be interested in verifying to what extent the aggregate predictive performance deteriorates as one neglects e.g. the stochastic dependence among the series in the model for sub-aggregates.

The problems of disaggregate vs. aggregate modeling and forecasting (also with focus on inflation forecasting) has been considered by many authors. In particular Hubrich (2005) considers similar inflation forecasting problem, but does not deal with density forecasts. At a more general level, the question was addressed by Hendry and Hubrich (2011). Some theoretical developments are related to the work of Giacomini and Granger (2004); Lütkepohl (2009) considers aggregation within an interesting class of disaggregate DGPs. Contributions involving also empirical applications, especially on inflation forecasting, include Aron and Muellbauer (2013), Castle and Hendry (2010) and Faust and Wright (2013) among others. The distinctive features of the approach taken here include the focus on density forecasting and the role of exclusion restrictions at the disaggregate level.

## 2. Model Framework

In order to establish the notation used below, some fairly standard results concerning estimation of multivariate linear models are recalled. Consider a Gaussian VAR( $p$ ) model for  $m$  variables:

$$\mathbf{y}_t = \boldsymbol{\alpha}^s + \mathbf{y}_{t-1}\mathbf{A}_1 + \dots + \mathbf{y}_{t-p}\mathbf{A}_p + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T, \quad (1)$$

where:

$\mathbf{y}_t$  is  $m$ -dimensional row vector (corresponding to period  $t$ ),

$\boldsymbol{\alpha}^s$  is  $m$ -vector of seasonal intercept terms (consisting of  $mS$  unknown parameters in total, where  $S$  is the number of different seasons)

$\mathbf{A}_1, \dots, \mathbf{A}_p$  are  $m \times m$  matrices of parameters, with  $\mathbf{A}_p$  including at least one non-zero element,

$\{\boldsymbol{\varepsilon}_t\}$  is a Gaussian vector white noise process with covariance matrix  $\boldsymbol{\Sigma}$ .

Assuming  $m = 1$  leads to a Gaussian autoregressive process AR( $p$ ). However, if one assumes that  $\mathbf{A}_1, \dots, \mathbf{A}_p$  and  $\boldsymbol{\Sigma}$  are diagonal, (1) describes a set of  $m$  unrelated Gaussian AR( $p$ )-type processes.

Assuming  $\Sigma$  to be positive definite and otherwise unrestricted and  $\mathbf{A}_1, \dots, \mathbf{A}_p$  to be unrestricted (i.e. containing no zero restrictions) results in a standard VAR formulation, where OLS estimates of the parameters are also MLEs and thus are asymptotically effective. However, for unrestricted  $\Sigma$  if  $\mathbf{A}_1, \dots, \mathbf{A}_p$  include zero restrictions (i.e. the set of explanatory variables is not the same in all the equations), OLS estimates are no longer equivalent to ML estimates (and are not asymptotically efficient anymore). Issues of non-stationarity (existence of unit or explosive roots in the characteristic polynomial of (1)) are not considered here, the autoregressive parameters are assumed to be unrestricted for the sake of simplicity. The initial conditions  $\mathbf{Y}_0$  are assumed to be fixed and the values are taken from the data, as the further analysis is conditional on  $\mathbf{Y}_0$ . In order to consider asymptotically efficient estimation of the model parameters in special cases (with exclusion restrictions), the following SUR-type form of (1) is considered:

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}, \tag{2}$$

where:

$\tilde{\mathbf{y}} = (\mathbf{y}_1 \ \mathbf{y}_2 \ \dots \ \mathbf{y}_T)'$  is a  $Tm$ -dimensional column vector,

$\boldsymbol{\beta}$  is a  $k$ -dimensional column vector consisting of unrestricted elements of  $\mathbf{A}_1, \dots, \mathbf{A}_p$  and  $\boldsymbol{\alpha}^1, \dots, \boldsymbol{\alpha}^s$ ,

$\tilde{\mathbf{X}}$  is  $Tm \times k$  matrix consisting of rows and columns arranged in a way that matches the convention assumed for  $\tilde{\mathbf{y}}$ ,

$\tilde{\boldsymbol{\varepsilon}} = (\boldsymbol{\varepsilon}_1 \ \boldsymbol{\varepsilon}_2 \ \dots \ \boldsymbol{\varepsilon}_T)'$  is  $Tm$ -dimensional column vector of Gaussian error terms with zero mean and covariance matrix of the form  $\mathbf{I} \otimes \Sigma$ .

Asymptotically efficient estimation of  $\boldsymbol{\beta}$  can be achieved by means of a Zellner-type estimator of the form:

$$\hat{\boldsymbol{\beta}}_{\Sigma} = [\tilde{\mathbf{X}}'(\mathbf{I} \otimes \Sigma^{-1})\tilde{\mathbf{X}}]^{-1} \tilde{\mathbf{X}}'(\mathbf{I} \otimes \Sigma^{-1})\tilde{\mathbf{y}}, \tag{3}$$

with  $\Sigma$  replaced by its consistent estimate,  $\mathbf{S}_{\boldsymbol{\beta}}$ , based on residuals (e.g. obtained in the previous iteration of the procedure, as the procedure could be iterated):

$$\mathbf{S}_{\boldsymbol{\beta}} = T^{-1}\mathbf{E}'\mathbf{E}, \tag{4}$$

where  $\mathbf{E}$  is  $T \times m$ -dimensional matrix of residuals. In order to initialize the procedure, the first estimate of  $\boldsymbol{\beta}$  can be obtained as  $\hat{\boldsymbol{\beta}}_{OLS} = (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}' \tilde{\mathbf{y}}$ . The procedure is assumed to stop once some estimate of  $\boldsymbol{\Sigma}$  is considered the final one, here denoted by  $\hat{\boldsymbol{\Sigma}}$ , with  $\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\Sigma}}}$  being the final estimate of  $\boldsymbol{\beta}$  (being also an approximation of MLEs for  $\boldsymbol{\beta}$ ).

### 3. Forecasting Methodology

For the purpose of one-period-ahead prediction, it is assumed that the point forecast of  $\mathbf{y}'_{T+1}$  (its estimated conditional expectation) is given by:

$$\hat{\mathbf{y}}'_{T+1} = \tilde{\mathbf{x}}^f \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\Sigma}}}, \quad (5)$$

where:

$\tilde{\mathbf{x}}^f$  is an  $m \times k$  matrix consisting of known constants and lagged values of the dependent variables (all these are readily available for one-step-ahead forecast), preserving the structure of  $\tilde{\mathbf{X}}$ .

For the sake of density forecasting, the predictive distribution of  $\mathbf{y}'_{T+1}$  considered here is  $m$ -dimensional Gaussian, with mean given by (5) and variance-covariance matrix given as:

$$\hat{V}_{T+1}^y = \hat{\boldsymbol{\Sigma}} + \tilde{\mathbf{x}}^f \left[ \tilde{\mathbf{X}}' (\mathbf{I} \otimes \hat{\boldsymbol{\Sigma}}^{-1}) \tilde{\mathbf{X}} \right]^{-1} \tilde{\mathbf{x}}'^f. \quad (6)$$

Consequently, a Gaussian distribution with mean and covariance matrix given by (5) and (6) respectively is perceived as an approximation to the Bayesian predictive distribution obtained with diffuse priors. The approximation can be quite a satisfactory one, especially if the number of observations is not very small and the prior information on parameters is not very strong.

A motivation for the use of the approximation instead of the exact Bayesian results is twofold. Firstly, the recursive of out-of-sample prediction experiment with extensive specification search imposes a considerable numerical burden, which requires some simplifications for the feasibility reasons. Secondly, one of the reasons often stated in favor of the Bayesian approach in VAR forecasting (e.g. with continuous Minnesota-type priors) is that the approach introduces some form of shrinkage, which is beneficial for the forecasting performance. In the paper the counterpart of shrinkage is the specification search described below. It could be perceived as analogous to

so-called Stochastic Search Variable Selection approach, which in turn imposes very strong (“hard”) shrinkage by means of discrete-continuous prior distributions. The paper focuses on the results obtained using the exclusion restrictions (similar to “hard” shrinkage) without imposing any “soft” restrictions (which usually imply reduction of the prior variance for certain parameters). The question whether imposing additional “soft” shrinkage could improve forecasting performance even more is left for further research.

It is assumed here that the variable of interest (labeled  $z_t$ ) is a linear combination of variables in  $\mathbf{y}_t$  with known, fixed weights  $\mathbf{c}_t$  :

$$z_t = \mathbf{c}_t \mathbf{y}_t', \tag{7}$$

where  $\mathbf{c}_t$  is a  $m$ -dimensional row vector of weights. Predictive distribution of  $z_{T+1}$  is therefore univariate Gaussian with mean given by:

$$\hat{z}_{T+1} = \mathbf{c}_{T+1} \tilde{\mathbf{x}}^f \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\Sigma}}}, \tag{8}$$

and variance of the form:

$$\hat{\text{var}}_{T+1}^z = \mathbf{c}_{T+1} \hat{\boldsymbol{\Sigma}} \mathbf{c}_{T+1}' + \mathbf{c}_{T+1} \tilde{\mathbf{x}}^f \left[ \tilde{\mathbf{X}}' (\mathbf{I} \otimes \hat{\boldsymbol{\Sigma}}^{-1}) \tilde{\mathbf{X}} \right]^{-1} \tilde{\mathbf{x}}^f \mathbf{c}_{T+1}'. \tag{9}$$

The above formulas simplify in certain cases, in particular when  $\boldsymbol{\Sigma}$  is assumed to be diagonal. The restriction leads to a considerable reduction of the numerical burden. Empirical verification of its predictive consequences is therefore of great practical importance.

It is obvious that *ex post* (point) forecast error of  $z_{T+1}$  is given by:

$$z_{T+1} - \hat{z}_{T+1} = \mathbf{c}_{T+1} \mathbf{y}_{T+1}' - \mathbf{c}_{T+1} \tilde{\mathbf{x}}^f \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\Sigma}}} = \mathbf{c}_{T+1} \left( \mathbf{y}_{T+1}' - \tilde{\mathbf{x}}^f \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\Sigma}}} \right), \tag{10}$$

so it is a linear combination of forecasts errors of  $\mathbf{y}_{T+1}$ .

Minimization of *ex ante* forecast error does not necessarily require the forecast error of  $\mathbf{y}_{T+1}$  to be minimized (in ‘mean squared’ sense for instance). In particular, it is possible that the disaggregate forecast errors just cancel out in the aggregate. Moreover, within a given model class, the forecast error resulting from application of the model directly to  $z_t$  could be higher compared to the one obtained for implicit forecast based on the analogous model for the disaggregate data  $\mathbf{y}_t$ .

As density forecasts are considered, it is of course crucial to provide adequate description of uncertainty: formulas (6) and (9) take into account potential stochastic dependence between the model equations (represented by estimated  $\Sigma$ ) and correlations within the joint distribution of the estimator for whole  $\beta$  in all the model equations (or a joint posterior for all the structural parameters). Therefore in general no ‘limited information’ approximation is used, unless it follows directly from the model structure (e.g. as a special case in which a full information method is equivalent to a limited information one). Another advantage of taking the general perspective of multivariate modeling is that it provides a framework in which the “naïve” disaggregate forecast strategy, based on separate sub-models, can be evaluated against more complicated alternatives.

The crucial issue is that of model choice (or comparison) at the disaggregate level. The approach used here is based on two basic premises. Firstly, importance of the exclusion restrictions is emphasized. This amounts to recognizing the fact that applying some reasonable variable selection or model selection procedures can lead to huge gains in predictive performance compared to a basic, unrestricted Gaussian VAR model. The related idea of Stochastic Search Variable Selection (see e.g. Frühwirth-Schnatter and Wagner, 2010) can be interpreted in terms of approximating the results of a Bayesian inference pooling experiment, dealing with the problem of finding the relevant model reducing restrictions. Introducing some form of parsimony is particularly important in the case of VAR models that are potentially heavily overparametrized.

Secondly, having in mind the model pooling or model comparison context, it seems to be reasonable to consider the criteria that measure predictive performance with respect to  $z_t$ . Consequently, though individual models are estimated on the disaggregate data, meaning that the estimation procedure itself aims at optimizing goodness of fit at the disaggregate level, the model comparison (or model selection) is undertaken based on predictive criteria at the aggregate level.

Similar issues to the ones presented above are formally considered by Lütkepohl (2009) within a broader model class (though Lütkepohl does not attempt to introduce parsimony and does not consider density forecasts).

In order to describe the above procedure in detail, an overview of some criteria for density forecast evaluation is provided below.

#### 4. Density Forecast Evaluation

The awareness of the necessity of more careful investigation of the forecast uncertainty has become widespread in the econometric literature (see e.g. Clark and Ravazzolo, 2015). The discussion on *ex ante* predictability of the recent economic crisis has also contributed to the recognition of importance of the probabilistic approach to forecasting. Tasks such as evaluation of probability of extreme events are no longer considered non-standard. However, the strand of research is deeply rooted in history of the statistical inference methods and some of the ideas can be traced back to classical texts. However, for the purpose of the paper just a brief review of the selected techniques is presented.

Standard ways of reporting *ex post* forecasting performance for point forecasts include considering mean forecast error, mean absolute forecast error or root mean squared forecast error (RMSFE). However, the concept of absolute forecast error can be generalized into a density context, leading to so-called continuous ranked probability score (CRPS). The generalization is based on the fact that the CRPS formula simplifies to that of absolute error if the predictive distribution is represented by point mass. The definition of CRPS, together with a closed-form analytical formula for the cases where predictive distribution is Gaussian is given by Gneiting and Raftery (2007, p. 367). Here the negatively-oriented version of CRPS is used (labeled CRPS\* by the authors), which takes positive values and directly generalizes absolute error. For two predictive distributions with location parameter corresponding exactly to the actual outturn, CRPS (intuitively, ‘a density forecast error’) will be higher for the density that is more dispersed. For the sake of *ex post* predictive performance analysis, the averaged CRPS is reported.

Another popular measure is so-called log score, being just a logarithm of the value of probability density function corresponding to the predictive distribution at the point being the actual outturn. Log-score has a Bayesian interpretation, because a sum of the one-period-ahead log-scores in a recursive predictive experiment on expanding sample can be perceived as related to so-called Predictive Bayes Factor, being a modification of the Bayes Factor, which in turn is a basic method of formal Bayesian model comparison. Relative to the CRPS measure, log-score (and the Bayes Factor) is considered to be sensitive to tail outcomes. Differences between (log of) values of Gaussian pdf say two, three or four standard deviations away from the mean do not increase proportionally, therefore in case of mis-predicted outturn log-score will discriminate models stronger compared to CRPS. Details and

more elaborate theoretical justifications of the above measures are summarized by Gneiting and Raftery (2007).

As the analysis conducted here is intended to provide an approximation to the results of Bayesian inference, the log-score will be used as the main criterion used for the model choice.

## 5. Model Comparison Framework

The model framework used here has the advantage of nesting a simple “practitioner’s approach” of the following form. Consider the problem of inflation forecasting. A simple practitioner’s solution would be to use disaggregate data on sub-indices and apply say  $AR(p)$  forecasting models to each of the individual series. The total CPI forecast would be then obtained making use of the fact that weights of the sub-aggregates in the total index are usually known. Such an approach accounts for heterogeneity to some extent, but neglects possible dependence. Individual predictive models are chosen based on say goodness of fit for the individual series. However, an advantage of the framework provided here is that it highlights the fact that the collection of such individual processes can be perceived as a single model for total CPI forecasting. From such a perspective the models for sub-aggregates should be evaluated jointly rather than separately. Consequently, neither of the sub-models should be evaluated without taking actual combination of the remaining ones into account. This of course results in a big increase in the number of specifications being considered.

The approach makes some generalizations of the “simple practitioner’s approach” readily available. For instance it is easy to impose stochastic linkages between equations by allowing the contemporaneous variance-covariance matrix to be non-diagonal. Another option would be to include lags of the other sub aggregates into the individual equations. The framework allows for considering of a broad menu of models that differ substantially. One end of the model spectrum would be the simplistic approach using individual processes (which is likely to be too restrictive), whereas the other end would correspond to an unrestricted VAR model (which is very far from being parsimonious). A solution proposed here is to explore options that are somewhere in between, meaning that these are less restrictive than the simple approach, but introduce “hard” shrinkage (parsimony by exclusion restrictions) compared to full, unrestricted VARs. This would correspond to including only some lags of some variables in the equations at the disaggregate level.

Alternative models (or sets of restrictions) should be compared by the predictive accuracy of the aggregate, here measured with log-score. The number of possible combinations of the restrictions is way too large to allow for any systematic specification search. Instead, a random-search algorithm is applied. Even if it is not capable of achieving the global optimum, it is interesting to check whether it can find a combination of restrictions that results in a model with predictive performance clearly dominating the simple special cases already mentioned. If such a specification can be found within reasonable computational time, the strategy would be empirically attractive and the issue should be investigated in more detail.

The objective of the paper is to present a strategy that is feasible for more than ten disaggregate variables. For each model (a set of exclusion restriction that actually corresponds to some particular form of  $\tilde{\mathbf{X}}$ ) a predictive experiment is conducted, with recursive model reestimation and one step ahead prediction (with the number of repetitions denoted by  $N$ , being also a number of one-step-ahead forecasts obtained from one specification). Results of the repeated out-of-sample predictive exercise are summarized (here by a sum of log-score values) and the best specification is retained. A modification of the algorithm that aims at inference pooling instead of model selection could be also considered, though the possibility is not explored here.

As the stochastic specification search requires thousands of model specifications to be checked, the computational burden is substantial. The computational limitations are the reason for which the attention here is restricted to simple Gaussian VAR models only (instead of say VARMA class members), for analogous applications see George, Sun and Ni (2008).

One more thing should be kept in mind: if  $N$  is not large, the specification search might result in overfitting issues. This might show a spurious predictive gains that are not necessarily observed out-of-sample, resulting from using the statistical noise to obtain perfect prediction in the verification period. In order to consider possible empirical consequences in depth, it would be necessary to add another data window over which performance is not maximized but just analyzed. However, for Polish macroeconomic time series the number of available observations is not large enough to do so. In order to avoid overfitting problems one might want to promote specification parsimony (imposing e.g. a limit on the number of non-zero autoregressive parameters, or using some penalty function).

## 6. Empirical Analysis: Nowcasting Polish CPI Inflation Rate

In order to illustrate practical applicability of the approach and possible gains in terms of forecast performance an empirical analysis of Polish inflation is provided<sup>1</sup>.

The analysis makes use of month-over-month CPI inflation rate in Polish economy for the period 2002M01–2015M08, with  $T = 164$ , (see Figure 1.). The last  $N = 44$  observations are treated as a verification window<sup>2</sup>. It is important to notice that the verification period is quite challenging, as it includes a period of deflation unprecedented in Polish economy.

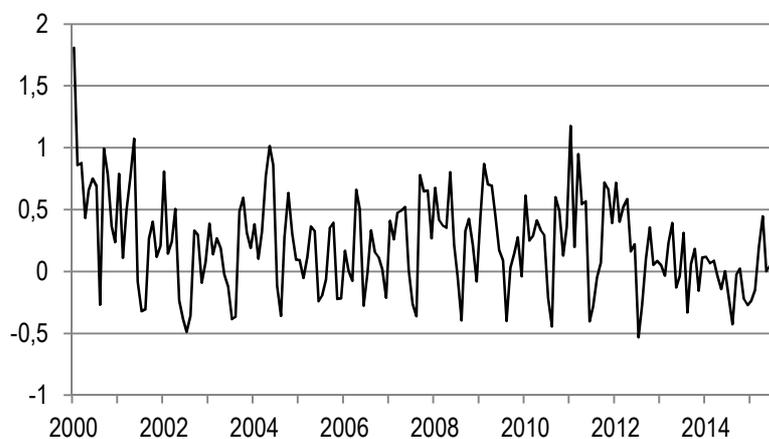


Figure 1. CPI m-o-m inflation rate [%] in Poland, 2000–2015

As for the data at the disaggregate level, Polish Statistical Office (GUS) reports disaggregation into 10 main price groups. However, as two of the components include both energy and non-energy prices, an effort has been made to separate the price indices. Moreover, housing expenditures and alcoholic beverages with tobacco are also separated from food, resulting in a total of 14 categories. The unrestricted model would therefore correspond to a VAR for 14 variables, having extra 196 parameters for each additional lag in the unrestricted version.

<sup>1</sup> Other studies examining similar problems include Clark (2006), Dees and Güntner (2014), Huwiler and Kaufmann (2013) and Ibarra (2012), see also Stock and Watson (2015).

<sup>2</sup> All the calculations were conducted using own routines written in Ox (see Doornik and Ooms, 2003), details of the specification search procedure are available from author by request.

The predictive exercise is pseudo-real time, which means that no revisions are taken into account (though for inflation the revisions are rather minor, occurring once a year only). It is assumed that there are no aggregation errors (that is that the percentage m-o-m change of the total CPI is equal to weighted sum of changes of the sub-aggregate categories), which is in accordance with (7). Moreover, the weights are assumed to be known, which is not always the case in real time, as the weights for a calendar year are published in March. The disaggregate categories are listed in Table 1, together with average weights in the sample period.

Table 1. Disaggregate CPI sub-categories used in the empirical analysis

No.	Code	Category label	Average weight
1	1a	Food and non-alcoholic beverages	0.258
2	1b	Alcoholic beverages, tobacco	0.060
3	2	Clothing and footwear	0.052
4	3a	Dwelling: housing, water (excluding energy)	0.116
5	3b	Dwelling: electricity, gas and other fuels	0.089
6	3c	Dwelling: Furnishings, household equipment and routine maintenance of the house	0.049
7	4	Health	0.050
8	5a	Transport: fuels for personal transport equipment	0.045
9	5b	Transport: excluding fuels	0.044
10	6	Communication	0.049
11	7	Recreation and culture	0.070
12	8	Education	0.013
13	9	Restaurants and hotels	0.052
14	10	Miscellaneous goods and services	0.053

*Note:* The average weight is computed as average over the years included in the sample (as the sample end does not correspond to December).

The benchmark models considered in the comparison are stationary and include unrestricted VAR models with one and two lags, as well as AR(1) process applied at the aggregate level, and a result of the specification search applied to AR( $p$ ) process for the aggregate data (resulting in omission of certain own lags of inflation). For the stochastic specification search algorithm, the maximum number of autoregressive parameters is limited to 42 (which reflects the idea that on average there are 3 parameters per equation, though these need not be uniformly allocated) and the maximum lag order is set to be equal to 13. Additional parameters (not included in the parameter count mentioned here) are the seasonal dummy variables, though the stochastic algorithm also switches between inclusion and exclusion of the seasonal dummies in each equation. Moreover, as for the contemporaneous covariance matrix two options are separately considered. In the first version,

the  $\Sigma$  matrix is assumed to be unrestricted (being of course positive-definite), in the other version it is assumed to be diagonal.

The results are reported in Table 2., including the number of parameters (with sub-total for intercept/seasonal terms). The criteria of forecasting performance include RMSFE and average CRPS (the lower the better) and summed up (decimal) log-score (the higher the better). All the criteria show consistent results (neither is contradicting the other ones).

Table 2. Comparison of the results from alternative specifications: predictive performance for aggregate inflation based on aggregate/disaggregate data

Model	$\Sigma$	st. dev.	RMSFE	CRPS	Log-score	No. of parameters	
						total	Intercepts
AR(1) for total CPI	–	0.248	0.250	0.142	–0.726	14	12
AR(2) for total CPI	–	0.249	0.249	0.142	–0.657	15	12
AR with variable selection	–						
total CPI		0.294	0.224	0.129	0.306	7	1
VAR(1)	nd	0.268	0.280	0.156	–2.283	455	168
VAR(2)	nd	0.280	0.348	0.180	–4.949	651	168
VAR with variable selection	d	0.228	0.174	0.100	5.058	147	91
VAR with variable selection	nd	0.247	0.187	0.107	3.668	224	102

*Note:* Full estimation results are available from the author by request. Verification period: 2012M01–2015M08; the forecasts are one-period-ahead (which amounts to nowcasting due to the publication lag). CRPS (Continuous Ranked Probability Score) is averaged over the realized forecasts, for log-score the sum of decimal logs is taken. For AR(1), AR(2), VAR(1), VAR(2) seasonal dummies are included. The second column indicates assumptions regarding the contemporaneous variance-covariance matrix (relevant for disaggregate models only), with ‘d’ denoting a diagonal matrix, whereas ‘nd’ denotes non-diagonal one; ‘st. dev.’ denotes *ex ante* forecast error (standard deviation of the predictive distribution) averaged over all forecasts. The column labeled ‘intercepts’ takes into account parameters corresponding to seasonality modelling as well.

Unrestricted VARs for disaggregate data show relatively poor performance, deteriorating with the number of lags. AR(1) and AR(2) models are similar, though AR with variable selection is clearly better. Poor performance of the AR(1), AR(2), VAR(1) and VAR(2) models results from restrictive assumptions about the highest lag order allowed, though VAR(2) is much worse than AR(2), so the other aspects seem to matter as well.

VARs with variable selection are clearly the best models, strongly dominating the other specifications by all the criteria. Moreover, the model assuming diagonal contemporaneous covariance matrix is getting slightly better results. In both cases *ex ante* errors are large compared to the RMSFE, consequently the forecast densities seems to be overdispersed.

One caveat should be mentioned here: as the models with non-diagonal covariance matrix of shocks and exclusion restrictions are computationally more demanding, the specification search algorithm might be less efficient

there. It is therefore possible that the algorithm used here gets trapped in a local optimum, which is not the global one, and the result could be somewhat improved upon a more intensive specification search. However, the practical conclusion would be that it is easier to find a useful model assuming uncorrelated shocks. Such a simplification is quite important from computational point of view as well.

The predictive gains obtained here are non-negligible, though the attention is restricted to one-period-ahead forecasts. However, the disaggregate information is likely to be useful in rather short horizons only. Moreover, it seems that the dependence between variables induced by incorporation of lags of other variables into equations is more important than the dependence represented by contemporaneous correlations of errors (at least from the viewpoint of quality of the forecasts of the aggregate series).

## Conclusions

The paper aims at demonstrating potential gains in terms of density forecasting that can be obtained by allowing for the use of disaggregate data. The analysis of density forecasts reflects the importance of the issue of uncertainty quantification in current econometric literature.

The explicit objective of forecasting of the aggregate series only is introduced here. Consequently, from a theoretical point of view it might be plausible to consider all the models for sub-aggregates jointly, as one model (even if the individual models are fully stochastically independent). The comparison of such models (or the corresponding sets of exclusion restrictions) is based on the predictive performance of the aggregate series.

The results obtained for Polish CPI inflation series confirm that considering multivariate models for the disaggregate data might result in substantial improvements in terms of prediction of the aggregate inflation. The conclusion remains true for various criteria for both point and density *ex post* forecast evaluation. However, the result stems from introducing parsimony-oriented exclusion restrictions. Some cross-variable dependence matters, as the best performing models include lags of other variables as well. On the other hand, there is no evidence that allowing for contemporaneous correlations of the shocks in disaggregate series is really important for the aggregate predictive performance in the case under consideration.

There are many possibilities to extend the scope of the analysis presented here, in particular by considering more general model classes. Besides that, two aspects are worth pointing out.

Firstly, a fully consistent theoretical framework for statistical inference in such situations should be developed. It was pointed out to the author that such a development could be based on a reparametrization of the observation space (at the disaggregate level) in such a way that the aggregate variable is explicitly considered<sup>3</sup>. This opens possibility for considering specifications taking advantage of concepts like the Bayesian Cut.

Secondly, the stochastic search algorithm used here to explore the model space (or, equivalently, the space containing various sets of exclusion restrictions) was rather simple and heuristic, with no theoretical premises for its efficiency. Perhaps some solutions used in the Stochastic Search Variable Selection setup could be adopted to match the specification search problem considered here.

The most important conclusion from the empirical analysis provided is that even within quite a simple class of multivariate models for disaggregate data a substantial improvement in predictive performance (over the standard unrestricted specifications) is possible. However, it requires a thorough specification search in terms of variable selection, which might be non-trivial when the dimension is large, which corresponds to detailed disaggregation.

## References

- Aron, J., Muellbauer, J. (2013), New Methods for Forecasting Inflation, Applied to the US\*, *Oxford Bulletin of Economics and Statistics* 75(5), 637–661, DOI: <http://dx.doi.org/10.1111/j.1468-0084.2012.00728.x>.
- Castle, J. L., Hendry, D. F. (2010), Nowcasting From Disaggregates in the Face of Location Shifts, *Journal of Forecasting*, 29(1–2), 200–214, DOI: <http://dx.doi.org/10.1002/for.1140>.
- Clark, T. (2006), Disaggregate Evidence on the Persistence of Consumer Price Inflation, *Journal of Applied Econometrics*, 21(5), 563–587, DOI: <http://dx.doi.org/10.1002/jae.859>.
- Clark, T., Ravazzolo, F. (2015), Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility, *Journal of Applied Econometrics*, 30(4), 551–575, DOI: <http://dx.doi.org/10.1002/jae.2379>.
- Dees, S., Guntner, J. (2014), Analysing and Forecasting Price Dynamics Across Euro Area Countries and Sectors: A Panel VAR Approach, *Economics Working Papers* 2014-10, Department of Economics, Johannes Kepler University Linz, Austria.
- Doornik, J. A., Ooms, M. (2003), Computational Aspects of Maximum Likelihood Estimation of Autoregressive Fractionally Integrated Moving Average Models, *Computational Statistics & Data Analysis*, 42(3), 333–348, DOI: [http://dx.doi.org/10.1016/S0167-9473\(02\)00212-8](http://dx.doi.org/10.1016/S0167-9473(02)00212-8).

---

<sup>3</sup> The author would like to thank Jacek Osiewalski for the valuable suggestion.

- Faust, J., Wright, J. H. (2013), *Forecasting Inflation*, in Elliott G., Timmermann. A. (eds.), *Handbook of Economic Forecasting*, vol. 2A, Amsterdam, North Holland, DOI: <http://dx.doi.org/10.1016/B978-0-444-53683-9.00001-3>.
- Frühwirth-Schnatter S., Wagner. H. (2010), Stochastic model specification search for Gaussian and partial non-Gaussian state space models, *Journal of Econometrics*, 154, 85–100, DOI: <http://dx.doi.org/10.1016/j.jeconom.2009.07.003>.
- George, E. I., Sun, D., Ni. S. (2008), Bayesian Stochastic Search for VAR Model Restrictions, *Journal of Econometrics*, 142(1), 553–580, DOI: <http://dx.doi.org/10.1016/j.jeconom.2007.08.017>.
- Giacomini, R., Granger, C. (2004), Aggregation of Space-Time Processes, *Journal of Econometrics*, 118(1-2), 7–26, DOI: [http://dx.doi.org/10.1016/S0304-4076\(03\)00132-5](http://dx.doi.org/10.1016/S0304-4076(03)00132-5).
- Gneiting, T., Raftery, A. (2007), Strictly Proper Scoring Rules, Prediction, and Estimation, *Journal of the American Statistical Association*, 102(477), 359–378, DOI: <http://dx.doi.org/10.1198/016214506000001437>.
- Hendry, D. F., Hubrich, K. (2011), Combining Disaggregate Forecasts or Combining Disaggregate Information to Forecast an Aggregate, *Journal of Business & Economic Statistics*, 29(2), 216–227, DOI: <http://dx.doi.org/10.1198/jbes.2009.07112>.
- Hubrich, K. (2005), Forecasting Euro Area Inflation: Does Aggregating Forecasts by HICP Component Improve Forecast Accuracy?, *International Journal of Forecasting*, 21(1), 119–136, DOI: <http://dx.doi.org/10.1016/j.ijforecast.2004.04.005>.
- Huwiler, M., Kaufmann, D. (2013), Combining Disaggregate Forecasts for Inflation: The SNB's ARIMA model, *Economic Studies* 2013-07, Swiss National Bank.
- Ibarra, R. (2012), Do Disaggregated CPI Data Improve the Accuracy of Inflation Forecasts?, *Economic Modelling*, 29(4), 1305–1313, DOI: <http://dx.doi.org/10.1016/j.econmod.2012.04.017>.
- Lütkepohl, H. (2009), Forecasting Aggregated Time Series Variables: A Survey, *Economics Working Papers* ECO2009/17, European University Institute.
- Stock, J. H., Watson, M. (2015), Core Inflation and Trend Inflation, *NBER Working Paper No. 21282*.

### Własności rozkładów predykcyjnych z modeli dla danych zdezagregowanych: prognoza inflacji w Polsce

**Z a r y s t r e ś c i.** W artykule podjęto kwestię weryfikacji występowania korzyści w zakresie poprawy jakości prognostycznej (oceniając także trafność *ex post* rozkładów prognoz) w przypadku prognozowania agregatu na podstawie modeli dla danych zdezagregowanych. Problem ten rozpatrywano w ramach modelu wektorowej autoregresji z restrykcjami, przy czym alternatywne specyfikacje odpowiadały różnym układom restrykcji zerowych. Zaprezentowano empiryczną analizę stopy inflacji CPI m/m w Polsce (rozpatrując 14 podkategorii dla danych zdezagregowanych). Modele z restrykcjami dla danych zdezagregowanych prowadziły do lepszych prognoz agregatu w porównaniu do modeli dla danych zagregowanych i modeli dla danych zdezagregowanych bez restrykcji (do porównania prognoz wykorzystano kryteria dla rozkładów prognoz takie jak CRPS oraz logarytm gęstości predyktywnej).

**S ł o w a k l u c z o w e:** predykcja, porównanie modeli, rozkład predykcyjny, inflacja, modele VAR

